Impact of Segments Aggregation on TCP Reno Flows in Optical Burst Switching Networks

Andrea Detti, Marco Listanti

***Abstract*-- In this paper we study the Optical Burst Switching (OBS) paradigm for the support of the TCP flows in an All Optical Network (AON). We analyze the TCP send rate, i.e. the amount of data sent per time unit, taking into account of: i) the burst assembly mechanism, called *burstification process*; ii) the burst loss events inside the OBS network. The goals of the paper are to investigate the effect of the variation of the *burstification* period and to derive some general guidelines about the dimensioning of the burstification period. With respect to the case in which any assembly mechanism is missing, the results show that an accurate dimensioning of the burstification period yields negligible penalties with regard to the low speed sources and significant benefits with regard to the high speed sources.**

A. INTRODUCTION

With the advent of the Wavelength Division Multiplexing (WDM) and with the rapid evolution and maturation of the optical technology, the *All Optical Networks* seem to be the candidate for the support of the future high speed IP backbone [1,2].

A sketch of a possible scenario for an optical IP network is depicted in Fig. 1. It consists in an WDM-based all optical backbone offering a transparent transport service to the adjoining electronic IP networks. The interface functions between the electronic and optical worlds are accomplished by the *Edge Nodes* (ENs), whereas the *Transit Nodes* (TNs) perform the switching functions exclusively in the optical domain.

It is foreseeable that in the near future an all optical backbone will offer high capacity circuit switched services by the provisioning of WDM end-to-end optical paths. In a longer term perspective a better use of the bandwidth will be attained by means of the optical packet switching.

An optical packet is thought to be composed of an header and of a payload. The header conveys the network layer control information allowing the TNs to perform the forwarding operation. Due to the absence of optical processing capability, the header is electronically processed, whereas the payload pass through the node directly in the optical domain.

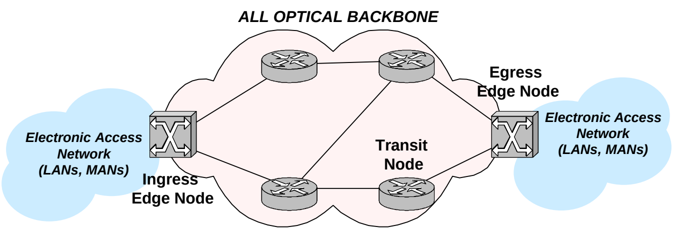

Fig. 1. All optical IP network scenario

A. Detti is with the Electronic Dept., University of Rome "Tor Vergata", Italy. E-mail: andrea.detti@uniroma2.it

M. Listanti is with the INFOCOM Dept., University of Rome "La Sapienza", Italy

The question to be solved is how to carry IP traffic via the optical packets. As each forwarding of an optical packet requires an electronic processing, in order to avoid that the processing load be the bottleneck of the network performance, it is desirable that the packet payload should be several times longer than the header. Moreover, the longer the optical packets are the higher the link efficiency is since the overhead due to the guard times between optical packets needed to cope with the configuration times of the optical devices can be neglected.

Unfortunately, a single IP packet is not so long to satisfy the previous requirement, so it is needed that several IP packets must be aggregated in a single optical packet and, consequently, it is required to implement the optical packet assembly and disassembly functions inside the ENs.

As far as the choice of the optical packet length is concerned, two solutions have been proposed: fixed size or variable size packets. The former is basically adopted in the *Optical Packet Switching* (OPS) [3,4,5], whereas the latter is utilized in the *Optical Burst Switching* (OBS) [6,7,8,9].

OPS is based on a synchronous node operation and on a coupled transport of header and payload. On the contrary, OBS allows an asynchronous node operation and uses a wavelength decoupling of the packet payload, named *Burst*, from its header, called *Burst Control Packet* (BCP). In this paper we basically refer to the OBS technique, even if the achievements can be also extended, at least qualitatively, to a OPS environment.

In the framework of the OBS technique, a link of the IP optical backbone supports $W+n$ wavelengths: W wavelengths, called *data wavelengths*, are dedicated to the burst transmission, whereas the remaining n, called *control wavelengths*, are signaling channels devoted to the transport of the BCPs. An ingress EN forms the bursts aggregating a number of IP packets directed towards the same egress EN. This operation is named *burstification* and is performed by a device called *burstifier*; accordingly, an ingress EN has to be equipped by as many burstifiers as the egress ENs are. Obviously, the burst must be structured in order to allow the receiving EN to properly delineate and extract each IP packet contained in the burst.

Once the burst is ready, the ingress EN sends the BCP aimed to reserve a free data wavelength on each link of the path. After an *offset time*, the EN injects the burst on the previously reserved optical virtual path. It is easy to recognize the reservation strategy closed to the well known *Tell and Go* [15,26].

As for the handling of the output contentions between burst at a TN, here we assume a bufferless node structure [7,17]. Burst contentions are handled in the wavelength domain by forwarding the conflicting bursts on different output wavelengths possibly with the use of Tunable Optical Wavelength Converters.

The previously illustrated issues related to an OBS network have been widely investigated in literature [10,11,12,13]; however, at the knowledge of the authors, contributions to the impact of the OBS mechanisms on the external tunneled protocols are not available. Nowadays, data communications are prevalently regulated by the TCP/IP protocol stack [14]. The IP protocol covers routing and forwarding functions, whereas, TCP assures a reliable end-to-end connection and adapts the data sent per unit time to the network conditions by means of the well known congestion control mechanisms.

From a general point of view, it can be argued that the burstification process can cause some *delay* penalties on the TCP flows. As a matter of fact, once reached an ingress EN, a TCP segment has to wait for the end of the burst aggregation time before that it can be forwarded, imbedded in the burst, towards the egress EN. This extra delay can determine a lowering of the bandwidth of the TCP connection. Moreover, the burstification process may introduce a level of *correlation* among the loss events of the TCP segments that may compromise the TCP recovery mechanisms. In fact, several consecutive segments of the same TCP connection may belong to the same burst; the loss of a burst yields a sequence of lost segments. Obviously, the correlation effect is more and more emphasized as the number of segments of the same source contained within a burst increases. This number depends on the relationship between the *burstification period* (i.e. the burst aggregation time) and the bandwidth via the TCP source reaches the EN through the access IP network.

In this paper we investigate the delay and the correlation effects introduced by the burstification process in an OBS network on a TCP Reno connection.

An analytical model for the evaluation of the TCP send rate, i.e. the segment sent per unit time over the OBS path, taking into account the presence of a burstifier is developed. The TCP send rates obtained in the presence and in the absence of the burstifier are compared. For the latter, we borrow the analytical model reported in [19]. The figure of merit used for the comparison is the ratio between the two send rates, called *burstification factor*. In this quantity we will distinguish the term related to the delay and the term related to the correlation effects; so the two effects will be separately analyzed and their sensitivity to the variation of the burstification period will be studied. The obtained results will allow us to define general criteria useful for the dimensioning of the burstification period.

The paper is organized as follows: in section B we present the network model and explain in detail the delay and the correlation effects. In section C the analytical evaluation these effects on the TCP send rate is carried out, while the section D is devoted to comparison of the TCP send rate in presence and in absence of the burstifier. In section E the main conclusion of the study are summarized.

B. NETWORK MODEL

We consider the TCP connection model reported in Fig. 2. The endpoints of the connection are named *source* and *receiver* and are supposed to implement TCP Reno version. The source transmits TCP segment and the receivers sends back the ACKs. The Ingress EN contains the burstifier, whereas the deburstification functions are performed by the Egress EN.

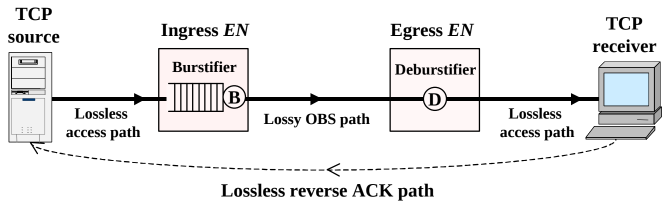

Fig. 2. TCP connection model

In the forward direction (i.e. from the TCP source towards the TCP receiver), we model the access network path as a lossless link with end-to-end delay equal to d and with bit rate equal to B_a bit/s (called *access bandwidth*). Moreover, the OBS path between the Ingress and the Egress EN is modeled as a lossy link with propagation time equal to T_p. The burst loss is Bernoulli distributed with parameter p. All the previously mentioned parameters (i.e. d, p, B_a, T_p) are considered to be constant.

In the reverse direction, we neglect the presence of the burstifier/deburstifier and we model this path as a lossless link with a fixed end-to-end delay equal to T_p+2d.

To simplify the model we assume that the transmission times of the TCP segments and of the OBS bursts are negligible as well as the delays due to the deburstification functions.

As far as the burstifier model is concerned, we refer to that proposed in [14]. In detail, the burstifier is modeled as a FIFO packet queue (Fig. 3) to which the TCP segments flow in. The queue is emptied (i.e. all packets are removed) after a constant time interval T_b, called *burstification period*, since the arrival of the first packet. The TCP segments enter the burstifier during a burstification period form the burst. We assume that the burst is emitted immediately as soon as it has been formed, so the effect of the *offset time* is not taken into account. However, it is possible to taking into account of the *offset time* value, increasing the T_p in the forward direction of the same value.

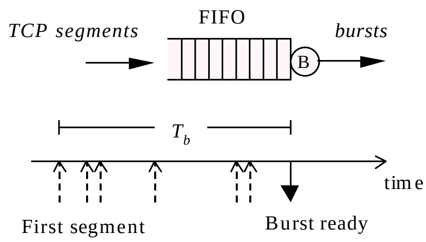

Fig. 3. burstifier logical sketch

The OBS network is assumed to be bufferless. Under this assumption, the relationship between the offered traffic and the burst loss probability, only depends on the amount of offered traffic (i.e. the *insensitivity property* [16,17]) and is independent of the burst length. From the previous reasoning, the value of T_b does not influence the burst loss probability p of the OBS network. Nevertheless, it is not painless with regard to the TCP performance.

To better explain the influence of the burstification period T_b on the TCP mechanisms, we distinguish three classes of TCP sources: *fast* , *medium* and *slow*.

A fast source has an access bandwidth (B_a) so high as to emit all the segments of its current congestion window (cwnd) within the interval T_b, so, an outgoing burst contains all the segments of its cwnd. On the contrary, a slow source has B_a so low as to emit at most one segment during T_b, therefore at most one segment of that TCP connection will be contained within an outgoing burst. The medium source has an intermediate behaviour. In formulas, fast, slow and medium sources satisfy the following conditions:

Fast sources $$\frac{W_m \cdot L}{B_a} \leq T_b \qquad (1)$$

Slow sources $$\frac{L}{B_a} \geq T_b \qquad (2)$$

Medium sources $$\frac{L}{B_a} < T_b < \frac{W_m \cdot L}{B_a} \qquad (3)$$

where:

- W_m : is the maximum cwnd advertised by the receiver at the connection establishment, measured in segments;
- L : is the segment size in bit;
- B_a : is the access bandwidth measured in bit/sec.

We expect that during a TCP connection the segments predominantly be of fixed length and so, we assume L as a constant.

A TCP segment is subjected to the delay due to the burstification process; in fact, it has to wait the expiration of the burstification period to be forwarded by the EN. This delay component increases both the end-to-end round trip time (RTT) and the retransmission time-out (RTO). Due to the TCP flow control mechanisms, it is straightforward to understand that the previous effects on the RTT and on the RTO let down the data rate of the TCP connection. We call this degradation as *delay penalties*.

Now, let us focus our attention, on the one hand, on what happens when a burst is lost along the OBS path and, on the other hand, on what happens when a burst is successfully delivered to the egress *EN*.

A TCP source belonging to the slow class experiences the loss of a single segment every time a burst loss takes place. As the burst loss events are statistically independent, the segment loss events are statistically independent as well; so, in the average a slow source experiences a segment loss every $1/p$ emitted segments (Fig. 4 left).

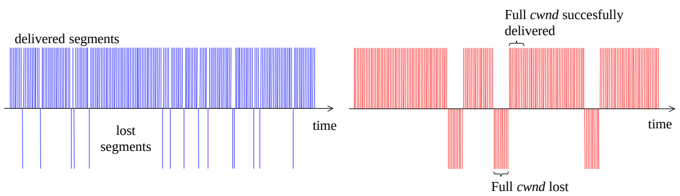

Fig. 4. Examples of *slow* and *fast* class lost and delivered segments traces.

On the contrary, a TCP source belonging to the fast class experiences the loss of all the segments of the current cwnd every time a burst is lost. Therefore, whereas the burst loss events are statistically independent, the segment loss events are highly time correlated. On the other hand, when a burst is successfully delivered to the egress EN, all the segments of the current *cwnd* are successfully delivered to the receiver. So, there is also an high correlation among the successful deliver events. In conclusion, a fast source experiences both "concentrated" losses and "concentrated" successful deliveries. In the average, one cwnd is completely lost every $(1/p - 1)$ cwnds successfully delivered (Fig. 4 right). Clearly, the fast recovery and fast retransmit recover mechanisms do not work for the fast sources, whereas they may be prevalent in the slow ones.

A medium source experiences segment loss events with a correlation level in the middle of the slow and the fast one. As well, the higher this level is, the nearer to the fast class boundary (1) the source is. In the next, we refer to these correlation effects as *correlation benefits*.

The Fig. 5 reports a typical trend of the TCP cwnd in two cases: i) the source is fast; ii) the source is slow. The losses of the fast source are recovered by means of the RTO mechanism; therefore, the cwnd falls down to one after each loss. Nevertheless, the concentrate successful deliveries, and consequently the concentrated ACKs reception, quickly reopens the cwnd so that it remains several times near to its maximum (e.g. W_m=128). On the contrary, the slow source losses are mainly recovered by means of fast recovery and fast retransmit mechanisms that do not throttle the cwnd as the RTO, but the shorter time intervals between two consecutive losses keep the cwnd significantly far from its maximum value.

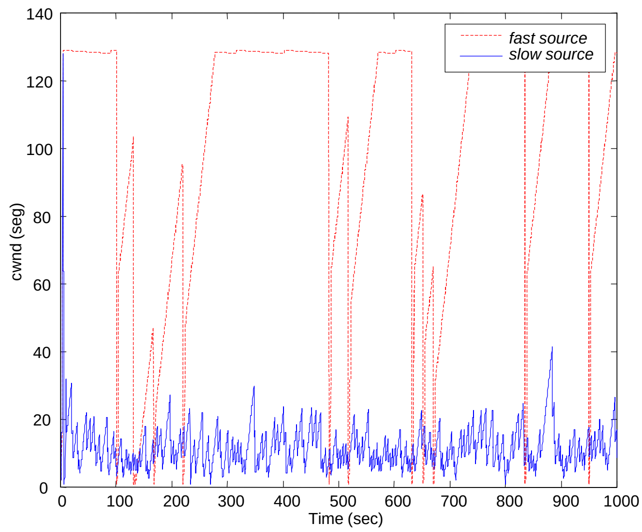

Fig. 5. TCP *cwnd* time trend

C. THE TCP RENO SEND RATE MODEL

In this section we develop an analytic model for the evaluation of the send rate of the TCP Reno. In the analysis we neglect the TCP timer granularity and do not model the delayed ACK feature [23], i.e. an ACK is sent for each data segment received.

In the same line of reasoning of [19], for any given time $t > 0$, let N_t be the number of segments sent in the time interval $[0,t]$, and $B_t = N_t / t$ (segment/sec) be the source send rate in that interval. Note that N_t is the number of emitted segments irrespectively their successful reception. We define the long-term steady-state *send rate* (B) of a TCP connection as:

$$B = \lim_{t \to \infty} B_t \tag{4}$$

In the rest of section, the increase of the RTT and the RTO due to the burstifier is firstly determined, then, the TCP send rate of the slow (B^s) and the fast (B^f) class is analytically evaluated. Finally we prove by means of a simulation approach that a source belonging to the medium class achieves a send rate (B^m) intermediate between the previous ones.

1st. Increase in RTT and RTO due to the delay penalties

Referring to the network model in Fig. 2, let us define:

- RTT : the average round trip time, which is the time period since the transmission of a segment to the reception of the related ACK;
- $RTTVAR$: the round trip time standard deviation;
- RTO^1 : the average value of the "first" retransmission time-out, which is the retransmission time-out without any *backoff* duplication [24].

The previous values include the contributions due to the presence of the burstifier. The following are the values of the same quantities in which the delay introduced by the burstifier is missing (i.e. in Fig. 2 the burstifier is absent). Hence,

- RTT_0 : the average round trip time in absence of the burstifier;
- $RTTVAR_0$: the round trip time standard deviation in absence of the burstifier;
- RTO_0^1 : the average value of the "first" retransmission time-out in absence of the burstifier.

Since in our network model (Fig. 2) the only delay variation is due to the burstifier, remembering the RTO evaluation rule of the TCP Reno [22], we have:

$$RTT_0 = 4d + 2Tp \tag{5}$$

$$RTTVAR_0 = 0 \tag{6}$$

$$RTO_0^1 = RTT_0 + 4RTTVAR_0 = RTT_0 \tag{7}$$

Let us define:

- $a = T_b / RTT_0$: the ratio between the burstification period and the round trip time in absence of the burstifier;
- $b = RTT / RTT_0$: the ratio between the round trip times with and without burstifier.

As the delay experienced by a segment within the burstifier is bounded in the interval $[0,T_b]$, the value of the RTT of every segment is overestimated by $RTT=RTT_0+T_b$, so an overestimate of b is given by $(1+a)$. Moreover, we assume $RTTVAR \approx RTTVAR_0$, that is the delay variation introduced by the burstifier is not so heavy as to significantly increase the round trip time standard deviation and, hence, the "first" retransmission time-out. Summarizing

$$b \simeq (1+a) \tag{8}$$

$$RTT \simeq (1+a)\,RTT_0 \tag{9}$$

$$RTO^1 \simeq (1+a)RTT_0 \tag{10}$$

The previous expressions indicate that the factor $(1+a)$ is the effect of the burstifier on the average round trip time and on the average "first" retransmission time-out.

2nd.Send rate for Slow class TCP sources

As previously mentioned, this class experiences independent segment loss events. So, our network model (Fig. 2) can be analyzed by means of the approach described in [19,20,21]. In particular, we utilize the send rate (expressed by (32) in [19]), here named B_{ku}, in order to derive the slow class TCP send rate (B^s). Hence:

$$B_{ku}(W_m, RTT, p, RTO^1) = \tag{11}$$

$$\begin{cases} \dfrac{\dfrac{1-p}{p} + E[W_u] + \widehat{Q}(E[W_u])\dfrac{1}{1-p}}{RTT(\dfrac{E[W_u]}{2}+1) + \widehat{Q}(E[W])RTO^1\dfrac{f(p)}{1-p}} & \text{for } E[W_u] < W_m \\ \dfrac{\dfrac{1-p}{p} + W_m + \widehat{Q}(W_m)\dfrac{1}{1-p}}{RTT(\dfrac{W_m}{8}+\dfrac{1-p}{pW_m}+2) + \widehat{Q}(W_m)RTO^1\dfrac{f(p)}{1-p}} & \text{othervise} \end{cases}$$

wherein,

$$E[W_u] = 1 + \sqrt{\frac{8(1+p)}{3p} + 1}$$

$$\widehat{Q}(u) = \min(1, \frac{3}{u})$$

$$f(p) = 1 + p + 2p^2 + 4p^3 + 8p^4 + 16p^5 + 32p^6$$

so, from (9), (10) and (11) the *slow* class TCP *send rate* (B^s) can be written as:

$$B^s = B_{ku}(W_m, RTT_0(1+a), p, RTT_0(1+a)) \tag{12}$$

3rd. Send rate for Fast class TCP sources

In this section we develop a model of the TCP congestion control and RTO recovery mechanism, that captures the correlation effects introduced by the burstifier on a fast source. The TCP behavior is modeled as a succession of "rounds". The generic j-th round starts with the transmission of W_j segments, where W_j is the current cwnd. Once all of segments of the current cwnd are sent, the next segment will be not transmitted until

- the first ACK is received for one of these W_j segments, or
- the retransmission time-out (RTO) expires.

The start of the transmission of the next segment determines the end of the j-th round and the begin of the (j+1)-th one.

Due to the condition (1), all of segments emitted in a round are contained in a single burst. As a consequence, when a burst loss takes place, then all of segments of the round are lost; we call this kind of rounds *lossy rounds*. On the contrary, when a burst is successfully delivered to the egress EN all of segments of the round reach the receiver; we call this kind of rounds *successful rounds*.

The succession of the rounds is formed by a sequence of successful rounds and by a sequence of lossy rounds. The two kinds of sequences alternate themselves in time.

We define the *Time Out Period* (*TOP*) as the time period comprising a sequence of successful rounds and the following sequence of lossy rounds.

Because the fast recovery and fast retransmit are not triggered in the fast class, at the begins of a *TOP*, the cwnd grows up in according to the TCP *slow start* [22]. If no loss occurs, the slow start phase is followed by the *congestion avoidance* one, in which the cwnd linearly grows up to its maximum value W_m. When a lossy round occurs, as soon as the $RTO=RTO^1$ expires, the source throttles its cwnd to one and begins to retransmit all of the segments of the last round. For each subsequent consecutive lossy rounds, the source doubles its *RTO* until 64 RTO^1.

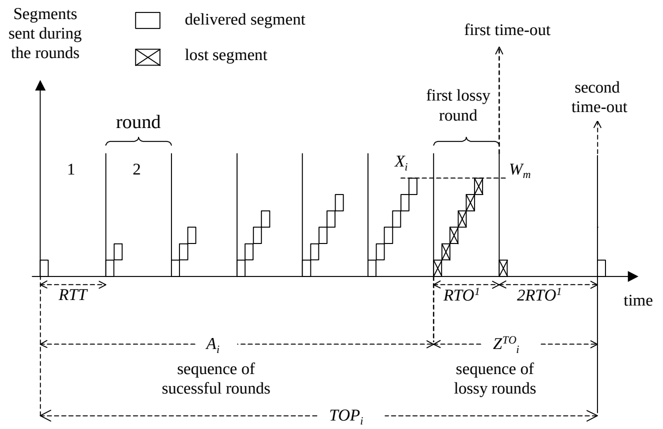

Fig. 6. Example of evolution of the *i-th TOP*

In Fig. 6 we show an example of the evolution of the cwnd during the generic i-th *TOP* assuming the slow start threshold *ssthresh* to be equal to one at the *TOP* beginning, i.e. the slow start phase is virtually missing.

At the first round the cwnd is equal to one and the source sends one segment (white box) that is successfully delivered to the receiver. After a time equal to *RTT*, the source receives the related ACK and the first round ends. For each subsequent successful round, the cwnd is incremented by 1, until it reaches its maximum value W_m. After X_i successful rounds, the segments sent in the (X_i+1)-th round are contained in a burst that is lost. As a consequence, all of segments are lost (crossed boxes). This is the first lossy round. After a time equal to RTO^1, the retransmission time-out (first time-out) expires. The cwnd is "throttled" to one and the first segment of the previous lossy round is retransmitted. This retransmitted segment belongs to a new round. This round is again a lossy round and a new time-out expiration event occurs after a time period equal to $2RTO^1$ (second time-out). Then, the source again retransmits the segment in a new round that is a successful round, hence, the i-th *TOP* ends.

For the generic i-th *TOP*, let us define:

- W_i : the value expressed in segments of the *cwnd* in the first lossy round;
- X_i : the number of successful rounds;
- R_i : the number of time-out expirations occurred in the i-th *TOP*;
- Y_i : the number of segments sent before the first time-out expiration;
- H_i : the number of segments sent after the first time-out expiration;
- A_i: the time duration of the sequence of the successful rounds;
- Z_i^{TO} : the time duration of the sequence of the lossy rounds;

We evaluate the average value of the TCP send rate of the fast class (B^f) as the ratio between the mean number of segments emitted in a *TOP* and its mean time duration, i.e.:

$$B^f = \frac{E[Y]+E[H]}{E[A]+E[Z^{TO}]} \tag{13}$$

Following a reasoning similar to that used in [19]:

$$E[Z^{TO}] = RTO^1 \frac{f(p)}{1-p} \tag{14}$$

where,

$$f(p) = 1 + p + 2p^2 + 4p^3 + 8p^4 + 16p^5 + 32p^6$$

$$E[R] = \frac{1}{1-p} \tag{15}$$

By observing the Fig. 6, it is possible to understand that the number of segments sent after the first time out is equal to the number of time-outs minus one. Therefore,

$$E[H] = E[R] - 1 = \frac{p}{1-p} \tag{16}$$

Due to the Bernoulli loss model assumption, the *X* random variable is geometrically distributed, i.e.

$$Pr\{X=k\} = (1-p)^k p \tag{17}$$

with average equal to:

$$E[X] = \sum_{k=0}^{\infty} k(1-p)^k \, p = \frac{1-p}{p} \tag{18}$$

Then we can compute the average duration of the sequence of the successful rounds

$$E[A] = E[X]\,RTT = \frac{1-p}{p} RTT \quad (19)$$

As far as the $E[Y]$ evaluation is concerned, we first consider two extreme cases: i) high burst loss probability and ii) low burst loss probability and the relevant values of $E[Y]$, i.e. $E^h[X]$ and $E^l[Y]$, respectively, are evaluated. Successively, we propose a general expression for $E[Y]$ and hence for B^f.

1) E[Y] for high burst loss probability: E^h

Here we suppose the loss probability to be so high as to assume that:

i) the cwnd limitations can be neglected, i.e. cwnd saturation is a quite rare event;
ii) *ssthresh*=1, i.e. the slow start mechanism does not operate and cwnd always linearly increases.

The assumed linear increase in the cwnd, allow us to say that:

$$W_i = X_i + 1 \quad (20)$$
$$Y_i = W_i (W_i + 1) / 2 \quad (21)$$

Hence, from the (17) we have:

$$Pr\{W=k\} = (1-p)^{k-1} p \quad (22)$$

$$E[W] = \sum_{k=1}^{\infty} k\left(1-p\right)^{k-1} p = \frac{1}{p} \quad (23)$$

$$E[W^2] = \sum_{k=1}^{\infty} k^2\left(1-p\right)^{k-1} p = \frac{2-3p+p^2}{(1-p)p^2} \quad (24)$$

using (21), (23) and (24), we are able to compute the E^h as follows,

$$E^h = \frac{E[W^2]}{2} + \frac{E[W]}{2} = \frac{1}{p^2} \quad (25)$$

2) E[Y] for low loss probability : E^l

Here we suppose the burst loss probability to be so low as to assume that:

i) $W_i = W_m$, i.e. before the first loss events the *cwnd* has already reached its maximum value;
ii) before the first loss events, the *cwnd* remains equal to W_m for a time much longer than the time required to reach W_m.

As a consequence of the previous assumptions, we can neglect the windows growth during the slow start and congestion avoidance phases and we can assume that at the start of the TOP the cwnd is in the W_m state. So:

$$Y_i = W_m\, X_i \quad (26)$$

$$E^l = E[Y] = \frac{W_m}{p} \quad (27)$$

3) E[Y] general expression

We observe that, if we use E^h for low loss probabilities, the assumption that cwnd is unconstrained makes the values of E^h greater than E^l, which, in this region of loss, is a tight model. Dually, in a region of high loss, it results $E^l > E^h$. For the previous reasons, we approximate $E[Y]$ as,

$$E[Y] \approx \min\left(E^h, E^l\right) = \begin{cases} E^h & \text{for } p > \dfrac{1}{W_m} \\ E^l & \text{otherwise} \end{cases} \quad (28)$$

Substituting (9), (10), (14), (16), (19),(28),in the (13) we have

$$B^f = \begin{cases} \dfrac{p^3 - p + 1}{(1+a)\,RTT_0\left[p(1-p)^2 + p^3 f(p)\right]} & \text{for } p > \dfrac{1}{W_m} \\ \dfrac{W_m - pW_m + p^2}{(1+a)\,RTT_0\left[(1-p)^2 + p^2 f(p)\right]} & \text{otherwise} \end{cases} \quad (29)$$

4th. Simulation study

In order to validate the previously proposed TCP models, the scenario in Fig. 2 has been simulated with the use of NS2 [25]. This section summarizes the results of the simulation study.

The reports the TCP send rate (bit/sec) obtained assuming T_b=3 ms, RTT_0=600 ms, W_m=128 and segment size L=512 byte. To better point out the difference among the source classes, we have simulated three access bandwidth scenarios: 1 Mb/s, 100 Mb/s and 200 Mb/s. These values are chosen so that, from (1) (2) and (3), the sources can be considered as slow, medium and fast, respectively.

Obviously [19], there is a close-fitting for the slow class model. About the fast class model, we note a “light” over estimation of the send rate around $p = 1/W_m$. As a matter of fact, this is the region of “medium” loss in which both the (25) and the (27) lightly overestimate *E[Y]*. Moreover, Fig. 7 confirms that the TCP send rate of a medium source (B^m) gets intermediate values between B^s and B^f, i.e.

$$B^s \le B^m \le B^f \quad (30)$$

As each source is subjected to the same delay penalties (i.e. the same T_b), the performance gap among the classes (Fig. 7) is only due to the different number of segments per burst. Then, we conclude that the higher the number of consecutive segments aggregated into a burst is, the higher is the send rate.

We stress that B^s and B^f are independent of the access bandwidth (B_a); whereas B^m depends of B_a.

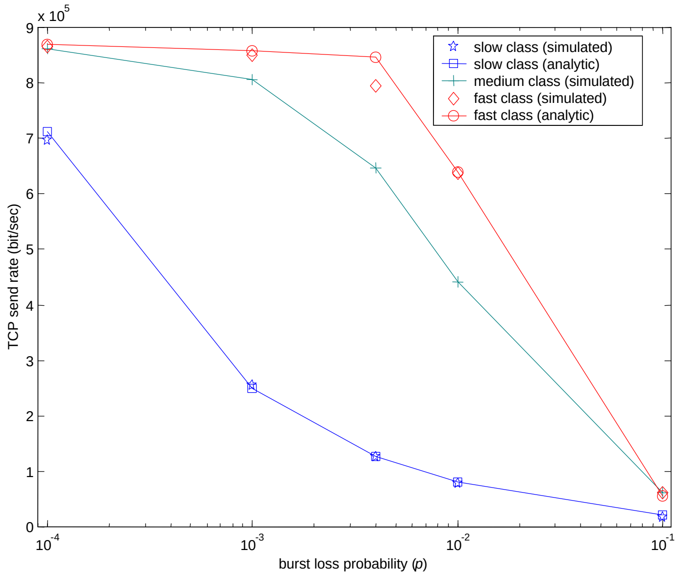

Fig. 7. TCP send rate vs. the burst loss probability (p) with T_b=3 ms, RTT_0=600 ms, W_m=128, L=512 byte, for several values of B_a, i.e. : 200 Mb/s (fast class); 100 Mb/s (medium class); 1 Mb/s (slow class).

D. Delay Penalties and Correlation Benefit Analysis

In this section, we compare the TCP send rates achieved in the presence and in the absence of the burstifier. This allows the *correlation benefit* and the *delay penalties* introduced by the burstifier to be clearly distinguished and quantified. Let us define:

- *NB* : the TCP send rate (measured in segment per second) achieved in the network scenario of Fig. 2 when the couple burstifier , deburstifier is missing;
- *B* : the TCP send rate (measured in segment per second) achieved in the network scenario of Fig. 2.

The figure of merit that is utilized for the comparison is the so called *burstification factor* (*F*), defined as the ratio between *B* and *NB*.

$$F = \frac{B}{NB} \tag{31}$$

This quantity aims at measuring the attenuation (i.e. F<1) or the amplification (i.e. F>1) of the send rate due to the presence of the burstifier.
If the burstifier is missing, the scenario of Fig. 2, is consistent with the hypotheses adopted in [19]. Moreover the round trip time and the retransmission time-out get the same value equal to RTT_0. Hence, we can use the (11) to calculate the TCP send rate in absence of the burstifier, i.e.

$$NB = B_{ku}(W_m, RTT_0, p, RTT_0) \tag{32}$$

In case the burstifier is present, as in the previous section, we determine the (31) distinguishing the source class. We analytically evaluate F of the slow class (F^s) and of the fast class (F^f). For (30) the burstification factor of a source belonging to the medium class (F^m) is in the middle of the previous ones.

$$F^s = \frac{B^s}{NB} \tag{33}$$

$$F^m = \frac{B^m}{NB} \tag{34}$$

$$F^f = \frac{B^f}{NB} \tag{35}$$

the (12) and the (29) can be easily expressed as,

$$B^s = \frac{B_0^s}{D_p} = \frac{NB}{D_p} \tag{36}$$

$$B^f = \frac{B_0^f}{D_p} \tag{37}$$

wherein ,

$$D_P = 1 + a = 1 + \frac{T_b}{RTT_0} \tag{38}$$

In (36) B_0^s represents the TCP send rate that may be obtained in absence of the burstifier (i.e. the (12) for $a = 0$). Hence, it is equal to *NB*.
In (37) B_0^f represents the (29) for $a = 0$. It is to be noted that, in this case, B_0^f can not be considered as the value of the send rate in absence of the burstifier, in fact, for $a \rightarrow 0$, the (29) does not hold any more since the assumption of fast source cannot be applied.
Substituting the (36) and the (37), in the (33) and in the (35), we have,

$$F^s = \frac{B_0^s}{D_p \ NB} = \frac{C_b^s}{D_p} \tag{39}$$

$$F^f = \frac{B_0^f}{D_p \ NB} = \frac{C_b^f}{D_p} \tag{40}$$

$$C_b^s = \frac{B_0^s}{NB} \tag{41}$$

$$C_b^f = \frac{B_0^f}{NB} \tag{42}$$

on the analogy of (39) and (40), let us define the (34) as,

$$F^m = \frac{C_b^m}{D_p} \tag{43}$$

and from (30) it results

$$C_b^s \le C_b^m \le C_b^f \tag{44}$$

In (39),(40) and (43) we have clearly distinguished the two burstification effects:

- D_p : measures the delay penalties;
- C_b^s : measures the slow class correlation benefit;
- C_b^f : measures the fast class correlation benefit;
- C_b^m : measures the correlation benefit of a source belonging to the medium class.

We stress that C_b^s and C_b^f are independents of the access bandwidth (B_a). On the contrary, C_b^m depends of B_a.
In the following we separately analyze the delay penalties and the correlation benefit. Afterward, we investigate on their joined action, i.e. on F, versus the access bandwidth and versus the burstification period.

1st. Delay Penalties
In a few words, the delay penalties reduce the send rate of a TCP Reno connection of a factor D_p, which is proportional to the ratio between the burstification period and the round trip time (without the burstifier).

2nd. Correlation Benefit
As expected, from the (36), the slow class correlation benefit is equal to one, i.e.

$$C_b^s = 1 \tag{45}$$

This means that a source belonging to the *slow class* does not experience any correlation benefit.
As far as the fast class correlation benefit is concerned, it is easy to prove the following properties of C_b^f :

i) it is independent both of RTT_0 and of T_b;
ii) it is equal to one in the extreme values of burst loss probabilities, i.e. p=0, p=1;
iii) it assumes its maximum for $p=1/W_m$. As a matter of fact, for p comprised in [0, $1/W_m$], the decrease rate (i.e. the derivative) of B^f is less than the NB one; whereas it is the contrary beyond $1/W_m$;
iv) the maximum value of C_b^f increases as W_m increases. This can be explained considering that the number of segments per burst increases;
v) let W_m^I and W_m^{II} be two values of the maximum cwnd (W_m), so that $W_m^{II} > W_m^I$. Let $C_b^f(W_m{}^I)$ and $C_b^f(W_m{}^{II})$ be the relevant values of C_b^f. For $p>1/W_m{}^I$, $C_b^f(W_m{}^I) = C_b^f(W_m{}^{II})$; this is due to the model assumptions that, for $p>1/W_m$, consider the cwnd how if it were unconstrained (for this loss probability, also the [19] model makes the same assumption). As a consequence, the numerator and denominator of the (42) beyond $1/W_m$ are independent of W_m.

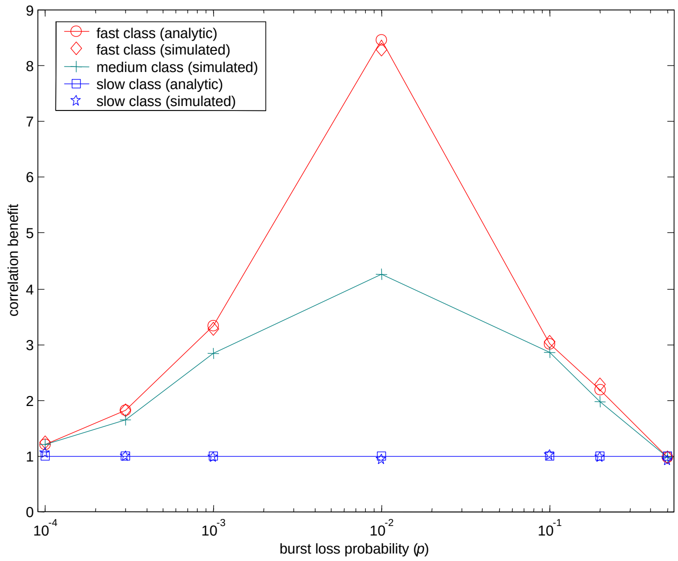

Fig. 8. C*orrelation benefit* vs. the burst loss probability (p) for T_b=60 ms, RTT_0=600 ms, W_m=128, L=512 byte, for several values of B_a, i.e. : 10 Mb/s (*fast class*); 3 Mb/s (*medium class*); 50 Kb/s (*slow class*).

In Fig. 8 we report the analytic and simulated correlation benefit curves versus the burst loss probabilities p. The plot helps us to outline the following conclusions:

i) the correlation benefit has a cusp centered in $1/W_m$;
ii) the higher the number of segments within the burst, the higher is the correlation benefit;
iii) the correlation benefit may give rise to a significant send rate amplification in the region of loss around $1/W_m$ (and this justify its name).

3rd. Burstification factor analysis
In this section we analyze the burstification factor (F) versus the source access bandwidth (B_a) and versus the burstification period (T_b). According to its definition, reported in (31), the burstification factor aims at measuring the attenuation (i.e. F<1) or the amplification (i.e. F>1) of the send rate due to the presence of the burstifier.
Fig. 9 shows both the simulated values and the theoretical results, i.e. (39) (40), referred to the burstification factor as a function of the access bandwidth, for burst loss probability equal to 10^{-2}, and for two values of burstification period: 60 ms (i.e. a=0.1) and 300 ms (i.e. a=0.5). As all the parameters of the (32) are constant, NB is constant, as well. Hence, an increase in the burstification factor means an effective increase in the TCP send rate (B).
Fixed the burstification period (e.g. T_b = 60ms), changing the access bandwidth (B_a) leads the following effects:

i) for B_a less than the slow class boundary (2), i.e. $B_a \le L / T_b$, the source puts only one segment within the burst, therefore, it does not get the correlation benefit, as shown in (45). As consequence, the burstification factor F is a constant equal to $1 / D_p$. It is worth to note that for very slow access bandwidth, F goes up again towards one. In fact, the TCP source begins to work without continuity solution (i.e. the time needed to send a cwnd is more than RTT) and the delay penalties are not able to interrupt this work modality. Hence, they do not worsen the send rate. The previous situation is not well modeled by the (11), that assumes the RTT to be greater than the time needed to send the cwnd [19]. In conclusion, the value $1 / D_p$ can be considered as a worst case of F ;

ii) for B_a beyond the fast class boundary (1), i.e. $B_a \geq (W_m L) / T_b$, the source puts the whole cwnd within the burst, therefore, it get the maximum gain from correlation benefit, i.e. C_b^f . As consequence, F is constant and equal to C_b^f / D_p .

iii) increasing B_a from the slow class boundary to the fast class boundary, the number of segments per burst increases. Hence, the correlation benefit and the burstification factor increase, as well.

In order to evaluate the effects due to the burstification period (T_b) change, let us consider two values of this parameter, T_b^I (e.g. 60ms) and T_b^{II} (e.g. 300ms), with $T_b^{II} > T_b^I$. On the access bandwidth plane (B_a), we define bs_{II} and bf_I, respectively, the slow class (2) boundary, for $T_b = T_b^{II}$, and the fast class (1) boundary, for $T_b = T_b^I$

$$bs_{II} = \frac{L}{T_b^{II}} \tag{46}$$

$$bf_I = \frac{W_m \; L}{T_b^I} \tag{47}$$

In general, when increasing the burstification period from T_b^I to T_b^{II} , the following consequences arise:

i) the *delay penalties* increase, as shown by (38);

ii) for $B_a \leq bs_{II}$ and $B_a \geq bf_I$, the number of segments per burst, and hence the correlation benefit, do not change; In fact, for both values of T_b , the number of segments per burst is equal to one for $B_a \leq bs_{II}$, and to the whole cwnd for $B_a \geq bf_I$;

iii) for $bs_{II} < B_a < bf_I$, the number of segments per burst, and hence the correlation benefit, increase.

The above consequences lead to the following effects in the burstification factor.

Obviously, if the correlation benefit does not increase, the increase of the burstification period from T_b^I to T_b^{II} decreases the burstification factor. In Fig. 9, this occurrence takes place for access bandwidth beyond 10 Mb/s ($B_a \geq bf_I$).

On the other hand, an increase in the correlation benefit can overcome the delay penalties increase. In Fig. 9, this is the case of access bandwidth in the interval [50Kb/s , 8Mb/s]. For B_a equals to 50 Kb/s, 5 Mb/s and 8 Mb/s, the *delay penalties* increase is predominant on the *correlation benefit* one. The opposite occurrence takes places for the other values of the interval.

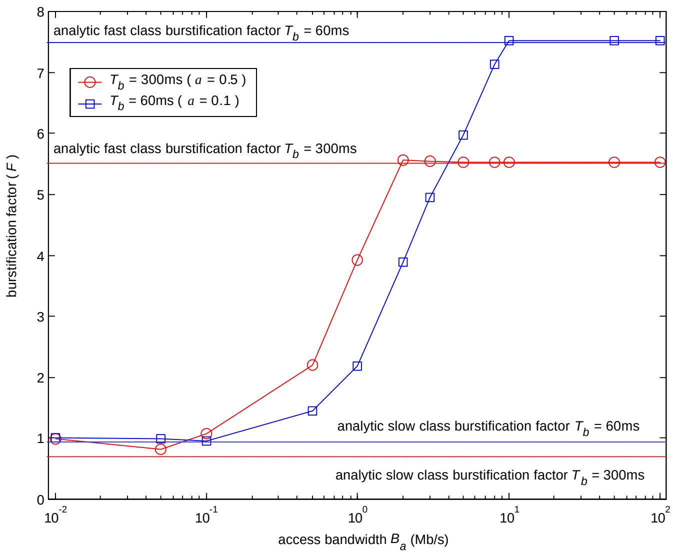

Fig. 9. Simulated B*urstification factor* (F) vs. access bandwidth (B_a) for RTT_0=600ms , W_m =128, L=512 bytes, p=10^{-2}

E. CONCLUSIONS

In this paper we have investigated the relationship between the burstification period and the TCP Reno send rate in an OBS IP optical network.

The analysis has outlined two opposite effects, namely the delay penalties and the correlation benefit.

The delay penalties are due to the delay experienced by the segments within the burstifier. They yield a send rate decrease proportional to the ratio between the burstification time and the round trip time evaluated in the case the burstifier is missing.

The correlation benefit regards the time correlation among the segment loss events and among the segment delivery events; i.e. due to the aggregation mechanism, a TCP connection may inserts a certain number of consecutive segments into the same outgoing burst; so, the burst loss/deliver event yields a consecutive segment loss/deliver events. The number of segments aggregated in the same burst depends on the relationship between the source access bandwidth and the burstification period. The obtained results have shown that, the more segments a connection aggregates inside a burst, the higher the correlation benefit is. Moreover, the correlation benefit is maximized for value of loss probability equals to the inverse of the maximum congestion window and it vanishes in the extreme values of loss probabilities.

According to the values of the access bandwidth value, the burstification period and the loss of the network, the correlation benefit may or may not overcome the delay penalties.

As far as the criteria of the burstification period are concerned, a reasonable choice seems to be around the 10% , 20% of RTT_0. In fact, the sources that reach the burstifier with small bandwidth, lightly worsen their send rate (with respect to the case of burstifier absence); whereas, those sources that have an high speed access experience an increase in the send rate that may be even very high.

In actual scenario, the value of RTT_0 is often unknown and can be different among the TCP connections that cross the burstifier. Instead, we can know the round trip time inside the bufferless OBS network; that is equal to twice time the end to end propagation time (RTT_{obs}) and is less than RTT_0. A prudential choice is to fix the burstification period equal to the 10% , 20% of RTT_{obs}.

REFERENCES

[1] B. Mukherjee, "Optical Communication Networks", McGraw-Hill Series on Computer Communications, 1997

[2] M. Listanti , V. Eramo, R. Sabella, "Architectural and technological issues for future optical Internet networks", *IEEE Communications Magazine*, Vol. 38, No. 9, September 2000, pp. 82-92

[3] S. Yao, B. Mukherjee, S. Dixit, "Advances in Photonic Packet Switching: An Overview", *IEEE Communications Magazine*, Vol. 38, No. 2, February 2000, pp. 84-94

[4] Callegati, M. Casoni, C. Raffaelli, B. Bostica "Packet Optical Networks for High-Speed TCP-IP Backbones", *IEEE Communications Magazine*, Vol. 37, No. 1, January 1999, pp. 124-129

[5] P. Gambini *et al.*, "Transparent Optical Packet Switching : Network Architectur and Demonstrators in the KEOPS Project", *IEEE Journal on Selected Area in Communications,* Vol. 16, No. 7, September 1998, pp. 1245 -1259

[6] C. Qiao, "Labeled optical burst switching for IP-over-WDM integration", *IEEE Communications Magazine*, Vol. 38, No. 9, September 2000, pp. 104-114

[7] C. Qiao, M. Yoo "A Novel Switching Paradigm for Buffer-less WDM Networks", *Proceedings of Optical Fiber Communication Conference (OFC)*, Paper ThM6, Feb. 1999, pp.177-179.

[8] Y. Chen, J. Turner "WDM Burst Switching for Petabit Capacity Routers", Proceedings of Milcom, 1999

[9] J. Turner, "Terabit Burst Switching", *Journal of High Speed Networks*, Vol.8, No.1, 1999, pp. 3-16

[10] C. Qiao, M. Yoo, "Choices, Features and Issues in Optical Burst Switching (OBS)", *Optical Networking Magazine*, Vol.2, April 1999.

[11] F. Callegati, A.C. Cankaya, Y. Xiong, M. Vandenhoute, "Design issues of optical IP routers for Internet backbone applications", *IEEE Communications Magazine* , Vol. 37, No. 12 , Dec. 1999, pp. 124-128

[12] J. Xiong, M.Vandenhoute, A.C. Cankaya, "Control Architecture in Optical Burst-Switched WDM Network", *IEEE Journal on Selected Areas in Communication*, Vol. 18, No. 10, October 2000

[13] M. Yoo, C. Qiao, S. Dixit, "QoS performance of optical burst switching in IP-over-WDM networks", *IEEE Journal on Selected Areas in Communications*, Vol. 18, No. 10, October 2000, pp. 2062-2071

[14] A. Ge, F. Callegati, L.S. Tamil, "On optical burst switching and self-similar traffic" *IEEE Communications Letters* , Vol. 4, No. 3, March 2000, pp. 98-100

[15] A. Detti, M. Listanti, "Application of Tell & Go and Tell & Wait Reservation Strategies in a Optical Burst Switching Network: a Performance Comparison", *Proceedings of IEEE International Conference on Telecommunication (ICT)*, Vol.2, pp. 540-548, June 2001

[16] F. P. Kelly, "Reversibility and Stochastic Networks", New York, Wiley,1980.

[17] A. Detti, V. Eramo, M. Listanti, "Performance Evaluation of a New Technique for IP Support in a WDM Optical Network: Optical Composite Burst Switching (OCBS)", submitted to *IEEE Globecom 2001*

[18] K. Thompson, G. J. Miller, R. Wilder, "Wide-Area Internet Traffic Patterns and Characteristics", *IEEE Network*, Vol. 11, No.6, pp. 10-23

[19] J. Padhye, V. Firoiu, D.F. Towsley, J.F. Kurose, " Modeling TCP Reno performance: a simple model and its empirical validation", *IEEE/ACM Transactions on Networking*, Vol.8, No. 2, April 2000, pp. 133 –145

[20] T. Lakshman, U. Madhow, "The performance of TCP/IP for network with high bandwidth-delay product", *IEEE/ACM Transactions on Networking*, Vol. 5, No. 3, June 1997, pp. 336-350

[21] M. Mathis, J. Semke, J. Mahdavi, T. Ott, "The macroscopic behavior of the TCP congestion avoidance algorithm", *ACM/SIGCOMM Comput. Commun. Rev.*, Vol 27, No. 3, July 1997

[22] W. Stevens, "TCP slow start, congestion avoidance, fast retransmit, and fast recovery algorithms", RFC2001, January 1997.

[23] W. Richard Stevens, " TCP/IP Illustrated,Volume 1: The Protocols", Addison-Wesley Professional Computing Series

[24] V. Paxson, M. Allman, " Computing TCP's Retransmission Timer", RFC2988, November 2000

[25] "Network Simulator 2" [OnLine] developed by Lawrence Berkeley Network laboratory and University of California Berkeley; http://www.isi.edu/nsnam/ns/

[26] John Y. Wei, Ray I. McFarland, "Just-In-Time Signaling for WDM Optical Burst Switching Networks", *IEEE Journal of Lightwave Technology*, Vol. 18, No. 12, December 2000, pp. 2019-2037