This is an author's version of the paper

Marchetto G.

"*High-Priority First Transmission to Efficiently Support Service Differentiation in Just-In-Time OBS Networks.*"

Published in

Journal of Optical Communications and Networking, vol. 2 n. 12, pp. 1031-1041

The final published version is accessible from here:

http://dx.doi.org/10.1364/JOCN.2.001031

High-Priority First Transmission
to Efficiently Support Service Differentiation
in Just-In-Time OBS Networks

Guido Marchetto

Abstract—**The provision of service differentiation is an important aspect that has to be considered for the definition of next-generation networks due to the high heterogeneity of the traffic that will dominate the networks in a near future. This is particularly important in the context of the optical burst switching, which is emerging as one of the strong candidate technologies for the next-generation optical Internet. Preemptive contention resolution schemes are very effective solutions for providing service differentiation in such networks, however they cannot be applied together with the just-in-time signaling protocol because of the great loss in efficiency in terms of wavelength utilization and maximum achieved throughput that results when the number of preemptions becomes too large. This paper presents a preemption based service differentiation solution that is suitable for the just-in-time optical burst switching paradigm as it keeps the preemption probability (i.e., the probability of observing a preemption when a contention occurs) low. Within the proposed technique, bursts are created at their ingress node and combined into chains, arranging them in order of decreasing priority. Then, a conventional preemption scheme is adopted at core nodes to handle contentions. An analytical study is presented and some traffic scenarios are analyzed by simulation to evaluate the performance of the proposed method.**

Index Terms—**Just-In-Time Optical Burst Switching, Optical Internet, Preemptive contention resolution, Service Differentiation.**

I. Introduction: Motivation and Related Work

THE traffic on the Internet continues to grow exponentially. This leads to an increasing demand for bandwidth that Internet Service Providers will have to satisfy by developing new network infrastructures able to support high-rate transmissions. This renewal process has to involve both the edge and the core of the network in order to provide an end-to-end broad-band service to all the network users. Optical Burst Switching (OBS) [1]-[8] is an emerging network core technology that can fulfill such requirements. It is therefore very attractive for use in the backbone of the future Internet.

OBS is based on the transport of IP packets over a Wavelength Division Multiplexing (WDM) [9][10] network. Packets entering the OBS network are assembled in bursts at their ingress node. Different bursts are created for different egress nodes. After an out-of-band control packet has

G. Marchetto is with the Dipartimento di Automatica e Informatica, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy (e-mail: guido.marchetto@polito.it; phone: +39 011 564 7094).

performed resource reservation (i.e., wavelength assignment and switching fabric configuration) at each WDM switch belonging to the burst path, each burst is transmitted through the network. If no resources are available, then the burst is dropped. This allows bursts to be transmitted over bufferless pre-allocated high-capacity channels, with low delays and zero jitters. The time that elapses from the start of the control packet to the instant at which the burst is actually sent through the network is called the *offset time*.

The Just-In-Time (JIT) [11]-[13] protocol is one of the signaling schemes defined for OBS networks. According to this protocol, resources for a particular burst are reserved at a node as soon as the corresponding control packet reaches the node, i.e., resources are reserved but not used during the offset time of a burst. Although other proposed signaling protocols provide more efficient wavelength utilization — e.g., Just-Enough-Time (JET) [14][15] and Horizon [4] operate resource reservation just before the burst arrives at the node — the JIT scheme has been assuming a role of primary importance thanks to its greater simplicity. In particular, one of the main outcomes of the study presented in [16] is that current trends concerning the development of optical switch and electronic hardware technologies are leading the simplicity of JIT to outweigh any performance benefits of Horizon and JET.

However, JIT-based OBS still has an important open issue concerning the support of service differentiation, which has to be a key aspect in the definition of next-generation IP networks. The current Internet does not provide more than a best-effort service, i.e., all packets get the same treatment and there are no guarantees concerning end-to-end delays and packet loss rate. However, the constantly increasing presence in the network of heterogeneous traffic flows (ranging from web sessions to interactive online games and critical communications) requires next-generation networks to be able to support different levels of Quality of Service. In the context of OBS, this essentially means to ensure that, in case of burst dropping for resource unavailability, higher priority packets are successfully transmitted with higher probability. One of the most effective known methods to perform such a service is the *contention resolution preemptive scheme*, proposed in [17] in the context of JET-based OBS networks. In this approach, a priority is assigned to each created burst. Packets are classified according to their service class at each ingress node and assembled in different bursts which assume the corresponding priorities. At each core node, whenever an incoming burst

contends for the same wavelength with a previously arrived burst, the contention is resolved according to these two rules: assuming n and m are the priorities of the incoming burst and the previously arrived burst respectively, (*i*) the incoming burst preempts the transmission of the previously arrived if $n > m$; (*ii*) the incoming burst is dropped if $n \leq m$. Since k-priority bursts contain only k-priority packets, this contention resolution scheme ensures that higher priority packets are always transmitted in case of contention. This scheme is effective when adopted in JET-based networks, where the achieved throughput is not affected by the number of preemptions [18], but a loss of efficiency raises if the JIT protocol is used. As better described in the next section and confirmed by our simulation results, preemptions imply a resource release and reservation overhead which reduces the average throughput (i.e., a large overall packet loss probability) when the number of preemptions becomes large. [18] also presents a probabilistic preemptive scheme, which could be adopted in JIT-based networks to reduce the number of preemptions and hence increase the average throughput. In this scheme, higher priority bursts can preempt lower priority ones only with a given probability. This reduces the overall packet loss probability, but also the effectiveness of the offered service differentiation as high priority bursts can be dropped in favor of low priority ones. An analytical study of these methods is provided in [19]. A similar technique has been studied in [20] which improves the wavelength utilization; however, it cannot avoid the bandwidth waste due to preemptions. Preemptive solutions can also be applied to implement bandwidth-allocation schemes [21][22][23], which provide service differentiation by reserving a different amount of bandwidth to each traffic class. A drawback of these solutions is the additional complexity at core nodes, due to the necessary introduction of proper scheduling algorithms and fiber delay lines (FDL).

Other solutions have also been proposed in the context of the JET signaling protocol. Burst segmentation [24] derives from the above presented methods and consists in assembling bursts with packets belonging to different service classes, arranging them in order from the highest to the lowest priority. When a contention occurs, the tail of the previously arrived burst is dropped. Since lower priority packets are arranged in the tail of bursts, this method achieves a low loss probability for high priority packets. However, this solution does not avoid the loss of efficiency of preemptive schemes when adopted with JIT, because a preemption still occurs whenever two bursts contend for resources. Another solution [25] consists in prioritizing bursts by assigning them a different offset time. In particular, an extra offset time is given to high priority bursts in order to properly delay them. This allows high priority traffic to have better performance than low priority one. However, in JIT-based networks, the extra offset time results in an increase of the waste initial time during which resources at a node are reserved but no transmissions are performed, thus further lowering the average throughput. Other proposed techniques (e.g., [26]) are based on intentionally dropping bursts in order to obtain a controllable

burst loss probability for different service classes. These solutions are applicable in JIT networks, but [21] shows that the resulting overall burst loss probability is higher than that obtained by preemptive schemes. This is due to the excessive dropping caused by this scheme, which hence also reduces the resource utilization.

The recently proposed *burst cluster transmission* [27] aims at improving the wavelength utilization by deploying a non-preemptive contention resolution scheme on network nodes. Service differentiation is provided by combining bursts into clusters at the network edge, arranging them from the lowest to the highest priority. This solution is suboptimal regarding both the achieved average throughput and the provided service differentiation because of the non-preemptive scheme applied to resolve contentions. In particular, there are two main issues that affect the burst cluster transmission effectiveness: (*i*) higher priority packets can be dropped even if they contend with lower priority ones and (*ii*) bursts can also be dropped when contending with burst clusters that will be dropped at other nodes.

This paper describes *High-Priority First (HPF) transmission* of bursts (originally presented in [28]), a service differentiation technique that improves burst cluster transmission. In fact, it provides high efficiency in service differentiation aware JIT-based OBS networks while maintaining the effectiveness of the previously described preemptive contention resolution scheme [17]. The proposed solution is based on the burst assembly algorithm proposed for the burst cluster transmission technique: bursts, assembled using a priority based classification of packets, are combined into chains for transmission. But here, unlike the burst cluster transmission scheme, the bursts are arranged in order from the highest to the lowest priority and nodes operate according to a preemption based contention resolution scheme. This particular burst transmission scheduling sensibly reduces both the number of contentions and the preemption probability in case of contention, achieving the above mentioned properties.

The rest of the paper is organized as follows. Section II focuses on the JIT protocol by discussing its operating principles and its problems when used with a preemptive contention resolution scheme. Section III describes the HPF transmission technique and presents some analytical results which quantify the reduction of the preemption probability obtained when this solution is adopted instead of the conventional preemptive scheme. Section IV provides an extensive simulation study which further validates the proposed method, and Section V concludes the paper.

II. JUST-IN-TIME PROTOCOL OVERVIEW

In OBS, the ingress node of a burst sends a control packet (the SETUP message) before burst transmission. Such packet performs resource reservation at each node belonging to the burst path. No acknowledgement for confirming the connection establishment is sent back to the ingress node, which can start to transmit the burst without waiting for any message. Wavelength reservation and switch configuration are performed as soon as the control packet arrives at a node, then

the packet is forwarded to the next node. Hence, since burst transmission has to be started only when resources have been (possibly) configured along the entire path, an initial transmission delay is necessary at the ingress node. Defining t_p as the control packet processing time at each node, and N as the number of hops from the ingress node to the egress node of a burst, the initial transmission delay T_d can be evaluated by $(N + 1)\, t_p$. Typically, $t_p = 1\ \mathrm{ms}$. Therefore, at each node there is an idle time during which resources are reserved but no transmissions are performed. The resulting bandwidth waste is however generally small if compared to the average burst size. Regarding the resource release, two methods have been proposed: (*i*) the *explicit release* method, which consists in sending a control packet (the RELEASE message) at the end of the burst transmission, and (*ii*) the *estimated release* method, where no more control packets are sent after the burst transmission and resource release is performed using burst size information that the ingress node has to put in the SETUP message.

When a preemptive contention resolution method is used to provide service differentiation, the SETUP message also contains the service class of the burst. When a SETUP message referred to a *n*-priority burst arrives at the generic core node and there are no available resources, a contention occurs. If n is higher than the priority m of a previously scheduled burst, the SETUP message preempts the transmission of such burst. However, the entire offset time has to elapse before the new burst arrives at the node. Thus, at each node there is an idle time during which resources are reserved but not utilized whenever a preemption occurs. This results in a swift rise of the wasted bandwidth when the number of preemptions becomes large. Furthermore, a resource release procedure based on explicit RELEASE messages has to be started by the node at which the preemption occurs. Depending on which algorithm is applied (one-way or two-way [29]), one or two RELEASE messages are generated, with different efficiency. However, in both cases, all nodes belonging to the path of the preempted burst are reached by a RELEASE message which allows them to release the resources that have been reserved for such burst. As propagation delays are not zero, several milliseconds can elapse before these resources can be reused, with consequent further bandwidth waste. The following section discusses the HPF transmission of bursts, which can sensibly reduce bandwidth waste by reducing the preemption probability.

III. HPF TRANSMISSION

A. Operating Principles

In a preemption based service differentiation solution, every burst that arrives at a node can potentially cause a preemption if its priority is higher than the actual lowest priority. In fact, in case of contention, an incoming burst always preempts the transmission of a lower priority one. HPF transmission of bursts aims at scheduling burst transmissions in such a way to reduce the probability for an incoming burst to contend with one at a lower priority. Thus, it aims at reducing the preemption probability when a preemptive contention resolution method is used.

HPF transmission operates at the network edge and consists in transmitting bursts consecutively, in order of decreasing priority, so that they appear as composed into chains. Two modules are necessary at the ingress node to perform HPF transmission: a *per-egress node burst assembler* and a *per-output port burst transmission scheduler*. Furthermore, a per-egress node oriented mixed time-length based burst assembly algorithm is used to generate bursts.

As previously seen, the *burst assembler* operates according to the algorithm presented in [27] for burst cluster transmission. A burst assembly module for each possible egress node is present at each edge node of the network. The module is composed by M buffers (let M be the number of supported service classes) where incoming packets are stored according to their service class. Bursts are therefore assembled with packets of the same service class and a priority k is assigned to bursts composed by *k*-priority packets. At this point, the per-destination burst assembly algorithm performs burst generation. Its operating principles are the same as the well-known Max-Time-Min-Max-Length based algorithm [30] developed for OBS networks, but here it is deployed in a per-egress node fashion. Max-time $T_{\max}$ and max-length $S_{\max}$ thresholds are related to the entire group of M buffers, and are defined so that *M* bursts are generated and passed to the output queues for transmission whenever:

1. $T = T_{\max}$, where T is a per-egress node timer restarted when a packet arrives at a node and assembling buffers related to its egress node are empty, or
2. $\sum_{k=0}^{M-1} S_k = S_{\max}$, where S_k is the current size of the *k*-priority burst.

The min-length threshold is instead referred to each single burst; if the size of a burst is under this threshold when $T = T_{\max}$ or $\sum_{k=0}^{M-1} S_k = S_{\max}$, the data size of the burst is increased to min-length with padding, as defined in [30]. This is necessary to guarantee a minimum burst length, which has to be defined according to the electronic processing speeds of SETUP messages, switching speed, and maximum size of a single IP packet, as described in [31].

Once the M bursts have been created, they have to be transmitted consecutively, as previously depicted. In order for the bursts to be sent out in such fashion, the *burst transmission scheduler* serves the output queues in a round-robin order, from the highest priority queue to the lowest priority one. As in the burst cluster scenario, control packets could be transmitted so that the wavelength is left idle only at the end of the entire burst chain. Thus, SETUP messages are sent in such a way bursts within a chain are transmitted consecutively. Furthermore, if explicit release is used, a RELEASE message is sent after the chain, while, in the case of estimated release, the RELEASE message is not sent and resources are released when the chain transmission finishes.
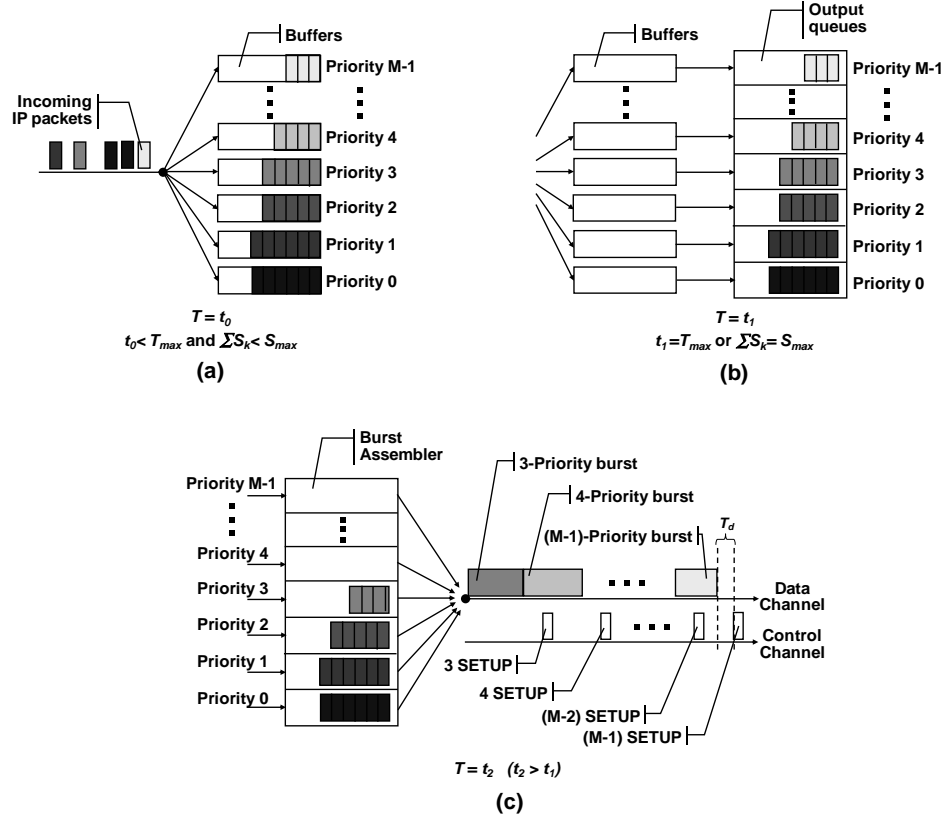
Fig. 1. HPF transmission operating principles: (a) packet arrivals and buffering, (b) burst assembly, and (c) burst transmission.

This transmission procedure introduces additional delays that are not present in traditional OBS networks. In fact, an entire chain has to be generated (i.e., $T_{\max}$ or $S_{\max}$ have to be reached) and then transmitted, thus causing both buffering and transmission delays. However, according to the principles of the Max-Time-Min-Max-Length based algorithm, the buffering delay can be controlled by properly defining the max-time threshold $T_{\max}$, while the maximum transmission delay is related to the max-length threshold $S_{\max}$. Hence, the max-time threshold $T_{\max}$ is defined according to the maximum tolerable assembling delay (some milliseconds are shown to be adequate in [27] and hence can be fine also in this case), while the max-length threshold $S_{\max}$ is fixed to $M \cdot S'_{\max}$, where $S'_{\max}$ is the maximum length threshold that would be used for a single burst in a conventional OBS network. This is done to avoid generating excessively long chains: the maximum amount of bytes sent within a chain composed by M bursts is equal to the maximum amount of bytes carried by M independent bursts in a traditional OBS network. Notice that the threshold $S_{\max}$ does not influence the delay of delay-sensitive packets (included in high priority bursts) as those are sent at the beginning of the chain. This is not the case in burst-cluster transmission, where bursts are transmitted in order of increasing priority and hence $S_{\max}$ has to be carefully defined to control the delay experienced by high priority packets, transmitted at the end of the chain [27]. In both cases, the maximum delay experienced by packets traversing the OBS network can be evaluated as

$$D_{\max} = T_{\max} + C \cdot S_{\max} + D_p, \qquad (1)$$

where C is the wavelength capacity and D_p is the propagation delay. This maximum value occurs when the max-length threshold is reached at the same time $T_{\max}$ expires.

Fig. 1 shows how the HPF transmission method operates in an edge node of an OBS network. Packet arrivals and buffering, burst assembly, and burst transmission in case of estimated release are presented. Notice that only $S_{\max}$ is used as max-length threshold in HPF. Hence, single bursts may exceed the $S'_{\max}$ value.

No modifications are required at core nodes, which operate according to the traditional preemptive contention resolution scheme [17].

B. Preemption probability evaluation

HPF transmission improves the efficiency of the conventional contention resolution scheme in JIT-based OBS networks by reducing the preemption probability. This section aims at giving an explicit quantification of this reduction, also providing an analytical comparison between the two methods.

The first thing to consider is the behavior of HPF when a contention occurs between two bursts, in particular when an incoming higher priority burst preempts the transmission of a lower priority one. The chain which the preempted burst belongs to is called the *preempted chain*. The other chain is the *preempting chain*. As the bursts within each chain are

transmitted consecutively, the preempting chain maintains the control of the wavelength until the last burst is sent. Furthermore, since bursts are transmitted in order of decreasing priority, the bursts that follow the preempted (preempting) one in the preempted (preempting) chain certainly have a lower priority. Hence, contentions without preemptions are observed when the remaining bursts of the preempted chain arrive at the node. In fact, they contend with higher priority bursts that belong to the preempting chain. Due to preemptions, chains are shortened at core nodes, i.e., highest priority bursts are transmitted while the remaining bursts are dropped in favor of other high priority bursts. Furthermore, they can also be split. This happens when a preempted chain acquires again a wavelength at some time after the preemption (e.g., if the preempting chain is short enough) and hence its lowest priority bursts are transmitted instead of being dropped. These events modify the shape of the chains flowing in the network. However, notice how these are not distorted (i.e., decreasing priority order is maintained for bursts). Thus, subsequent core nodes can also benefit from HPF transmission of bursts because, in case of preemption, the preempting chain still maintains the control of the wavelength until the last burst of the (shortened or split) chain is sent.

Fig. 2 shows an example of contention resolution in a congested node when HPF transmission is used at the network edge and when a conventional preemptive scheme is adopted instead. For the sake of simplicity, control packets are omitted and only one wavelength per port is shown. A service differentiation scheme based on $M = 8$ service classes (priority 7 is the highest) is used. Fig. 2(a) shows the case of HPF transmission. The contention between two complete chains is described for simplicity. Burst chain A arrives first at the node and acquires the output wavelength 3. The first burst of chain B arrives during the 5-priority burst transmission. Since its priority is higher than 5, it preempts the transmission of the chain A: 5-priority burst tail is dropped and burst chain B assumes the control of the wavelength. The following bursts of chain A contend with higher priority bursts of chain B and are therefore entirely dropped. In the conventional scheme depicted in Fig. 2(b), bursts arrive disordered at the node and hence several contentions are observed. Concerning HPF transmission, notice that even if chain A is cut during the contention resolution procedure, it is not distorted. This allows also subsequent core nodes to benefit from HPF transmission of bursts.

This example shows how HPF transmission can reduce the preemption probability with respect to the conventional scheme, where a preemption can potentially occur whenever two bursts contend. However, the algorithms adopted in HPF for assembling and transmitting bursts also have direct influence in the overall number of contentions at a node. In fact, this ordered accommodation of bursts results in a wavelength occupation which may be sensibly different than that obtained when the conventional scheme is adopted.

In order to obtain a quantitative evaluation of the benefits provided by HPF transmission in terms of achieved preemption probability, we developed a queuing model of the output port of an OBS node, considering both HPF and the conventional scheme. In both cases the node is supposed to be wavelength conversion capable. Furthermore, we introduced some simplifications. First, a single output port of a single OBS node is considered, as the analytical study of an entire OBS network would be extremely more complex and difficult to tract. Second, contentions among complete chains are considered in the case of HPF to simplify the analysis. Third, the HPF formulation disregards the possibility that a chain arrives at the node when a highest priority transmission is still ongoing and hence the incoming burst either is dropped or acquires another wavelength. The real preemption probability may be therefore larger than that evaluated by this model. However, with a reasonable number of service classes M and packet priority distribution, this simplification becomes less stringent as the probability for two high priority bursts to contend is reduced.
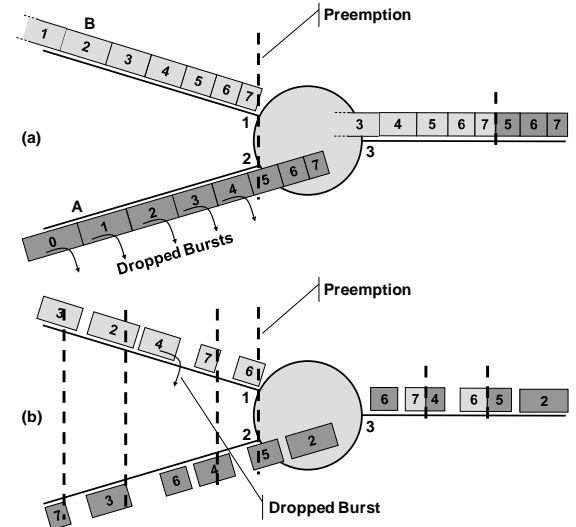

Fig. 2. Contention resolution: (a) HPF transmission; (b) Conventional scheme.

Despite these simplifications, the model provides an analytical support to the validation of the proposed approach, which however is integrated with an extensive simulation study where these simplifications do not hold. Simulation results will be presented in the next section. A complete analytical study considering an entire OBS network (and hence different burst and chain arrival patterns and lengths) is however interesting to further investigate the behavior of HPF. Hence, this will be subject of future work and publications.

TABLE I
SUMMARY OF USED VARIABLES
(CONVENTIONAL PREEMPTIVE SCHEME ANALYSIS)

Variable	Description
λ_k	k-priority burst arrival rate
μ_k	k-priority burst service rate
n_k	Number of k-priority bursts in the system
W	Number of wavelengths per link
M	Number of service classes
P_ξ	Probability for the system to be in state ξ
B_k	Set of the probabilities for the system to be

	in a state preemptable by a k-priority burst
P_{PR}	Preemption probability

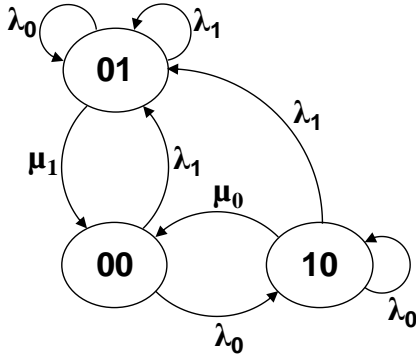| Variable	Description
λ'	Burst chain arrival rate
μ'	Burst chain service rate
ρ	Traffic intensity
C	Chain of bursts
b_k	k-priority burst in the chain
W	Number of wavelengths per link
M	Number of service classes
P'_{PR}	Preemption probability

Fig. 3. State transition diagram of an OBS output port adopting a conventional preemptive contention resolution scheme ($W=1$, $M=2$).

We start our analysis from the evaluation of the preemption probability achieved by the conventional preemptive contention resolution scheme. If for a given class of service k we assume Poisson distributed burst arrivals with mean rate λ_k and exponentially distributed service times (i.e., the burst lengths) with mean service rate μ_k, a conventional preemption based JIT output port can be modeled as a continuous time Markov chain. If W denotes the number of wavelengths per link and M the number of service classes, the generic state of the Markov chain can be defined as $\xi = (n_0,...,n_k,...,n_{M-1})$, $\sum_{k=0}^{M-1} n_k \leq W$, where n_k identifies the number of k-priority bursts currently in the system (see Table I for a summary of the notation adopted in this first part of analysis). The transition diagram modeling a simple output port where $W=1$ and $M=2$ is shown in Fig. 3. This simple configuration contains three states: (0,0), where no bursts are in the system, (0,1), where a 1-priority (the highest) burst is in transmission over the only available wavelength, and (1,0), where the wavelength is used by a 0-priority (the lowest) burst. State transitions occur when the system is idle and a new burst arrive, when a burst is finished to be serviced (i.e., to be transmitted over the wavelength), and when a burst is in the system but a new higher-priority arrival occurs and force a preemption of the channel (see Fig. 3).

The above defined Markov chain is homogeneous and aperiodic. Furthermore, it is irreducible as we can reach any

state from any other state. Hence, we can conclude that the chain is ergodic and a steady-state solution exists. If we define P_ξ as the probability for the system to be in the generic state $\xi = (n_0,...,n_{M-1})$, this can be evaluated for the simple configuration presented in Fig. 3 by solving the following system of equations:

$$\begin{cases} \lambda_1 P_{00} + \lambda_1 P_{10} = \mu_1 P_{01} \\ \lambda_0 P_{00} = \mu_0 P_{10} + \lambda_1 P_{10} \\ P_{00} + P_{01} + P_{10} = 1. \end{cases} \quad (2)$$

In particular, from (2) we can obtain the following expressions for the state probabilities:

$$P_{00} = \frac{\mu_1}{\lambda_0 + \lambda_1 + \mu_1}$$

$$P_{01} = \frac{\lambda_1 \cdot \ \lambda_0 + \lambda_1 + \mu_0}{\lambda_1 + \mu_0 \ \cdot \ \lambda_0 + \lambda_1 + \mu_1}$$

$$P_{10} = \frac{\lambda_0 \mu_1}{\lambda_1 + \mu_0 \ \cdot \ \lambda_0 + \lambda_1 + \mu_1}.$$

A preemption occurs when the transmission of a h-priority burst needs to be scheduled *iff* the system is in a state $\xi' = \ n'_0,...,n'_{M-1} : \sum_{k=0}^{M-1} n'_k = W$ and $\exists\, k < h : n'_k \neq 0$, $n'_k \in \xi'$. In the single-channel example represented in Fig. 3, a preemption occurs whenever a burst arrives at the system and the transmission of a lower priority burst is ongoing. Thereby, the overall preemption probability can be evaluated as:

$$P_{PR} = \frac{\lambda_1}{\lambda_0 + \lambda_1} P_{10}. \quad (3)$$

Given the set $B_h = \ P_\xi : \sum_{k=0}^{M-1} n_k = W, \sum_{i=0}^{h-1} n_i > 0$, i.e., the set of the probabilities of being in a state that is preemptable by an incoming h-priority burst, (3) can be extended in order to obtain the overall preemption probability of the conventional preemptive scheme presented in [17], for any possible values of W and M:

$$P_{PR} = \frac{\sum_{k=0}^{M-1} \sum_{P_\xi \in B_k} \lambda_k \cdot P_\xi}{\sum_{k=0}^{M-1} \lambda_k}. \quad (4)$$

All possible B_k, and consequently P_{PR}, can be evaluated numerically.

Let us consider now the HPF approach. In this scheme bursts are transmitted in chains. We assume Poisson distributed burst chain arrivals with mean rate λ'. Furthermore, as said above, we assume chains to be complete. In particular, let b_k denote a k-priority burst and $C = \ b_0,...,b_{M-1}$ a chain containing M bursts transmitted in decreasing order of priority (see Table II for a summary of the notation adopted in this second part of analysis). Given a chain C_a currently in transmission over an output wavelength and an incoming chain C_b, a preemption occurs when the

burst $b_{M-1} \in C_b$ (i.e., the first burst of the incoming chain C_b) contends with a burst $b_k \in C_a$, $k < M - 1$. This means that the preemption occurs if C_b arrives at the output port when the transmission of $b_{M-1} \in C_a$ is already finished. Under the above described assumptions, we can evaluate the preemption probability of an OBS output port operating according to the HPF scheme as the blocking probability achieved by a non-preemptive M/G/k/k queuing system where $k = W$ and service times are equal to the summation of the service times of the $M - 1$ lowest priority bursts, i.e., $1/\mu' = \sum_{k=0}^{M-2} 1/\mu_k$. This can be calculated by the well-known Erlang B formula, hence:

$$P'_{PR} = \frac{\rho^W / W!}{\sum_{i=0}^{W} \rho^i / i!}, \qquad (5)$$

where $\rho = \lambda' / \mu'$.

The conventional preemptive scheme [17] and HPF transmission — i.e., equations (4) and (5) — are compared in Fig. 4 and Fig. 5 for $M = 4$ service classes and different values of available wavelengths per link, i.e., $W = 16$ and $W = 24$, respectively. $\lambda_k = \lambda$, $\mu_k = \mu$, $\forall k : 0 \le k < M$ is assumed for simplicity, while $\lambda' = 1/M \cdot \sum_{k=0}^{M-1} \lambda_k$ is assumed in order to compare the two approaches under the same traffic load conditions. The analysis is also validated by simulation. The figures show how HPF transmission achieves a lower preemption probability than the conventional scheme for any intensity of offered traffic. The reduction is higher (more than one order of magnitude) for low values of traffic intensity, but it is however consistent also when the output port is close to saturation. As said, this property has a direct influence on the utilization of the network resources: a reduction of the number of preemptions also reduces the amount of bandwidth waste which occurs in JIT-based networks before the transmission of each burst, thus maintaining the efficiency of the network high notwithstanding the preemptive nature of the adopted contention resolution scheme.

C. HPF vs. Burst cluster

As described above, HPF transmission inherits some principles from the burst cluster transmission technique. However, both the effectiveness and the wavelength utilization achieved by the HPF approach are higher than that obtained with burst cluster transmission. In burst cluster transmission, incoming bursts are always dropped if there are no available resources. Since resources at a node could be reserved by bursts that will be subsequently dropped along their path, it is possible that incoming bursts are blocked in favor of bursts that never reach their destination. This effect, common to all non-preemptive schemes [32], results in a bandwidth waste that could rapidly increase at high traffic rate. We refer to this phenomenon as *bandwidth waste blocking*. HPF transmission can limit the effects of bandwidth waste blocking thanks to both the arrangement of bursts into the chains and the preemptive contention resolution scheme adopted by network nodes: only the low priority tails of the

burst chains could be blocked with consequent bandwidth waste. Furthermore, HPF transmission also avoids drops of high priority bursts that contend with lower priority ones, which are inevitable in a non-preemptive scheme such as burst cluster transmission.
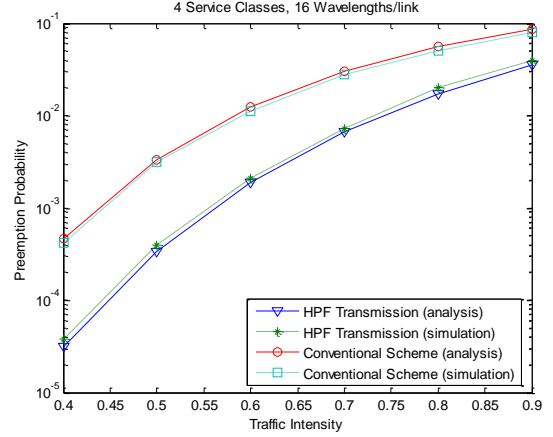

Fig. 4. Preemption probability at a single OBS output port ($W = 16$, $M = 4$).
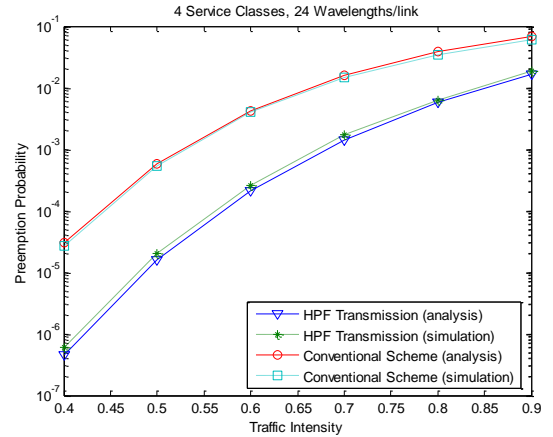

Fig. 5. Preemption probability at a single OBS output port ($W = 24$, $M = 4$).

IV. SIMULATION RESULTS

The analysis presented in the previous section helps in quantifying the gain of HPF transmission with respect to the conventional preemptive scheme in terms of preemption probability reduction at a single OBS node. However, in order to provide a complete evaluation of the proposed method, it is necessary to consider an entire optical network, where chains arriving at nodes may be incomplete (e.g., shortened or split) and hence may follow different arrival patterns. These are not covered in the above analytical model, which hence requires to be supported by further investigations. Furthermore, the negative effects of preemptions are more relevant when considering an entire network and really affect the overall packet loss probability, as explained in Section II. Hence, a simulation study is conducted over the 14-node NSF network topology showed in Fig. 6. These simulations compare HPF with the conventional preemptive resolution scheme, as well

as with the burst cluster transmission method, from which the burst assembly algorithm adopted in HPF is derived and which represents the state-of-the-art solution for the provision of service differentiation in JIT-based OBS networks. We consider both the overall packet loss probability and the achieved throughput.
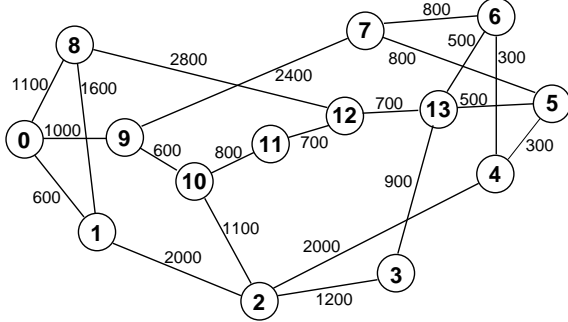

Fig. 6. 14-node NSFNET.

Packet arrivals are modeled by a Poisson process with average arrival rate λ. Furthermore, a given discrete probability distribution function models the belonging of packets to the defined service classes. In particular, an incoming packet has priority k with probability p_k. Ingress and egress nodes of an incoming packet are uniformly distributed among the 14 nodes composing the network topology. For simplicity, a fixed packet length of 1 KB is used. In addition, the number of wavelengths per link is $W = 8$, the capacity of each wavelength is $C = 10\,\mathrm{Gb/s}$, and the control packet processing time t_p is set to 1 ms. For the burst assembly algorithm parameters, max-time threshold $T_{\max}$ is set to 5 ms, while the per-burst max-length threshold $S'_{\max}$ is assumed to be equal to 7 MB. However, some experiments are run with different values to investigate the effect of these parameters on the system performance.

Network nodes are supposed to be wavelength conversion capable. For the sake of simplicity, a wavelength for preemption is randomly selected among all the wavelengths used by lower priority bursts [29]. Furthermore, the estimated release method is implemented.

Fig. 7 compares the conventional preemptive contention resolution scheme and the HPF transmission method with regards to the achieved packet loss probability when $M = 4$ service classes are used. Then, $S_{\max} = 28$ MB in this case. The packet priority probability distribution function is $p_3 = 0.1$, $p_2 = 0.2$, $p_1 = 0.3$, $p_0 = 0.4$. Priority 3 is the highest. It can be observed that both systems can effectively serve highest priority packets (packet loss probability achieved to 3-priority packets is low and comparable for both cases). However, if HPF transmission is not used, the great number of preemptions that occur at core nodes causes a gradual loss in efficiency that results in an unacceptable quality of service provided to lower priority packets. In fact, a high overall loss probability can be observed in Fig. 7 for lower priority packets, which results in an undifferentiated service offered to them at high packet arrival rate.

Fig. 8 shows how HPF transmission is also superior to burst cluster transmission in providing service differentiation. It is

noticeable that the dotted lines, related to burst cluster transmission, are closer to each other and higher than the lines that represent HPF transmission. This results from the preemptive nature of the adopted contention resolution scheme, which enables HPF transmission to (*i*) guarantee high priority packets to be transmitted when contending with lower priority ones, and (*ii*) reduce bandwidth waste blocking.

Fig. 9 plots the overall achieved throughput in the three considered cases and in a conventional preemptive JET network. The JET protocol achieves the best results, as expected, due to the limited effects that preemptions have on these networks [18]. However, it is noticeable that the average throughput offered by HPF transmission is higher than in the other considered JIT-based techniques and comparable to that of the JET scheme for low traffic load conditions. Fig. 9 also shows the effect of the bandwidth waste blocking (presented in Section III.C), which causes the average throughput of burst cluster transmission to be also lower than that of the conventional preemptive scheme at high arrival rate. In fact, when the network is highly loaded, the number of contentions increases, and consequently also the probability for a burst to be dropped grows. Since preemptions do not exist in burst cluster transmission, a burst is likely to be blocked by another which then is dropped along its path, thus sensibly lowering the average throughput. This result allows us to conclude that, under high traffic load conditions, the negative effect of preemptions in JIT-based networks is less significant than the bandwidth waste blocking in non-preemptive schemes.

In order to further investigate the efficiency of HPF, we evaluated the percentage loss concerning the resource utilization of both this technique and the conventional preemptive JIT scheme (where bursts are not transmitted in chains) with respect to the preemptive JET approach. The resource utilization is defined as the amount of time during which resources are reserved and actually used for transmission. Table III reports the results. Notice how HPF experiences only a 10.46% reduction in resource utilization with respect to the JET approach, while the conventional JIT scheme experiences a 35.76% reduction. This results in about a 39.41% gain in resource utilization of HPF with respect to the conventional JIT scheme.
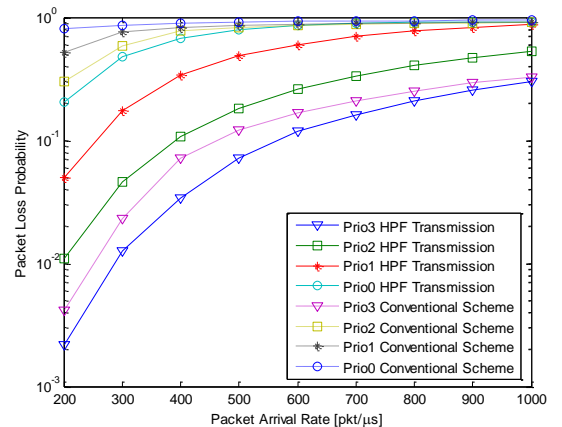

Fig. 7. Packet loss probability versus packet arrival rate for $M = 4$: comparison between HPF transmission and conventional contention resolution scheme.
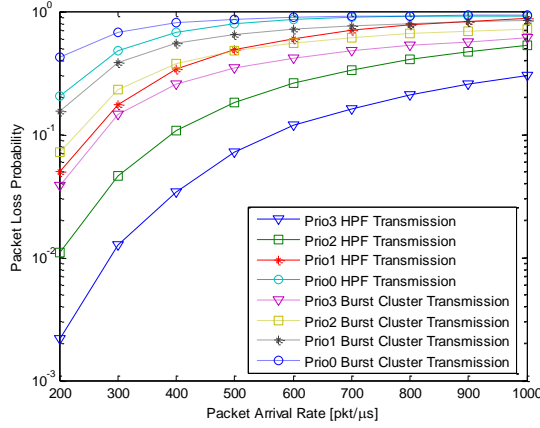
Fig. 8. Packet loss probability versus packet arrival rate for $M = 4$: comparison between HPF transmission and burst cluster transmission.
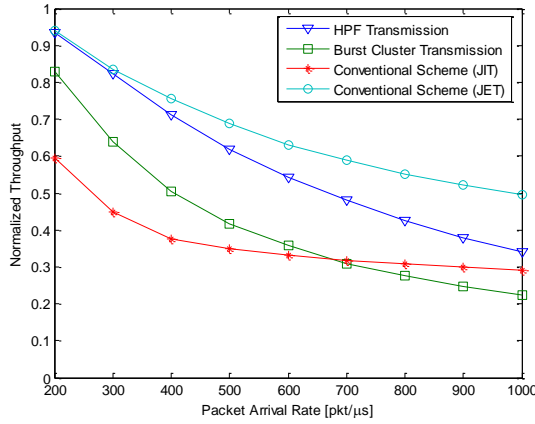

Fig. 9. Average throughput versus packet arrival rate.

TABLE III
PERCENTAGE LOSS IN RESOURCE UTILIZATION

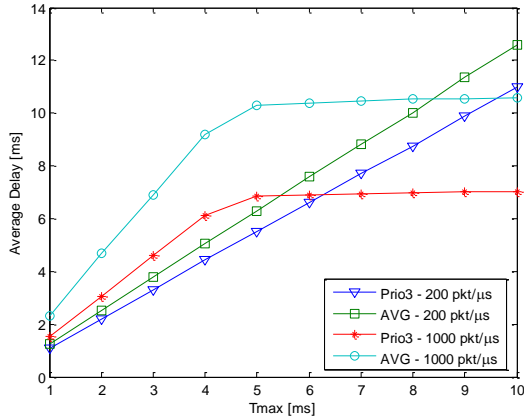| Scheme	Percentage Loss
HPF	10.46 %
Conventional Scheme	35.76 %

Fig. 10. Delay experienced at the entrance of the OBS network versus max-time threshold $T_{\max}$.

In HPF, bursts are transmitted in order of decreasing priority. As described in Section III.A, this ensures delay-sensitive traffic not to experience excessive delay due to both buffering and transmission over the wavelength at the entrance of the OBS network. This delay is mainly influenced by the max-time threshold $T_{\max}$ value. This considered, Fig. 10 plots the average delay experienced by packets at the entrance of the network at two different traffic load conditions, as well as the average delay of the highest priority (i.e., delay-sensitive) traffic. Notice the effect of $T_{\max}$ at low traffic load conditions, which limits the buffering delay as forces the chain to be transmitted after the assembly timer expiration. This effect is noticeable also at high traffic load when $T_{\max}$ assumes small values, but disappears when the max-time threshold value increases: when $T_{\max} > 4\,\mathrm{ms}$, the chain size reaches the maximum value $S_{\max}$ (which is fixed to $28\,\mathrm{MB}$ in these tests) within the max-time threshold and hence the chain transmission starts before the timer expiration. Thanks to this mixed time-size chain control, the delay of highest priority packets (as well as the average delay) is controllable and may be forced not to exceed given bounds.

While $T_{\max}$ is important to control the delay of delay-sensitive data, the max-length threshold $S_{\max}$ has a major impact on the maximum achievable throughput. This is common in JIT-based OBS networks as longer bursts reduce the overhead due to the initial offset time, during which resources are unused. Fig. 11 shows the offered average throughput as a function of the max-length threshold $S_{\max}$ in some traffic load conditions. The max-time threshold $T_{\max}$ is set to $5\,\mathrm{ms}$ in these tests. At low incoming traffic rate, the average throughput is high and almost constant as the network is underloaded and chains are generally created due to timer expiration, thus maintaining about the same length regardless of the $S_{\max}$ value. When the traffic increases, the average throughput clearly decreases, as also observed in Fig. 9. Moreover, low $S_{\max}$ values lead to a further reduced network throughput because shorter chains are transmitted as their size reaches the max-length threshold. Large threshold values would increase the throughput, but to the detriment of the buffering and transmission delay, which for this reason is controlled by properly setting $T_{\max}$. In fact, notice how also at high traffic rate the throughput remains almost constant for large $S_{\max}$ values due to the action of the max-time threshold $T_{\max}$.

The performance of HPF transmission is evaluated also with a higher percentage of high priority traffic. In particular, Fig. 12 considers the case of $M = 4$ uniformly distributed service classes and shows how HPF transmission can efficiently provide service differentiation also with such traffic scenario. However, it is noticeable that the high priority packet loss probability increases. In fact, when the amount of high priority packet increases, the probability for a packet to contend with another of the same (of higher) priority also grows, thus causing an increment of that priority packet loss probability. This considered, we investigated the effect of the percentage of highest priority packets on the loss probability of this type of traffic. In particular, Fig. 13 plots the highest

priority packet loss probability as a function of the probability for an incoming packet to belong to the highest priority service class. We fixed $p_3 = 0.1 \div 1$, with $p_i = (1 - p_3) / 3$, $i \neq 3$. Some traffic load scenarios are considered. We can observe how the 3-priority packet loss probability increases when p_3 grows. However, it tends to a constant value, which represents the performance of a non-preemptive network as when all packets have the same priority preemptions are no longer possible.
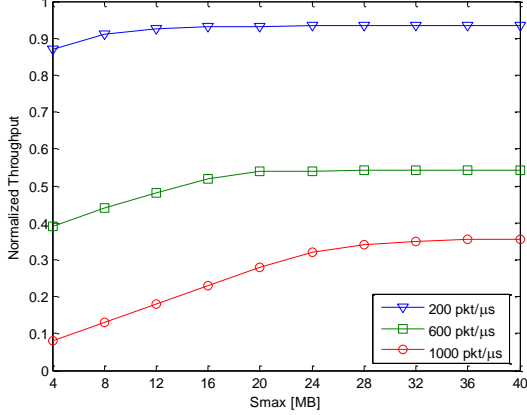
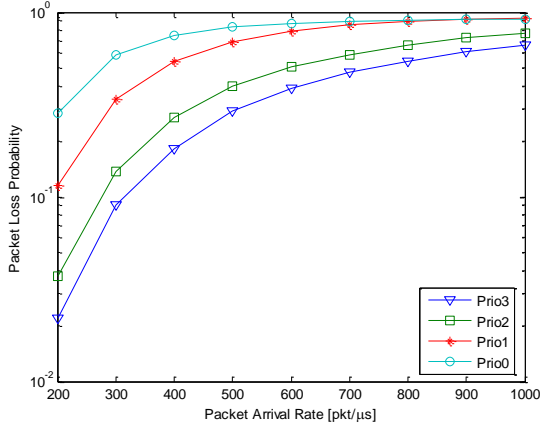Fig. 11. Average throughput versus max-length threshold $S_{\max}$.

Fig. 12. Packet loss probability versus packet arrival rate for uniformly distributed service classes.
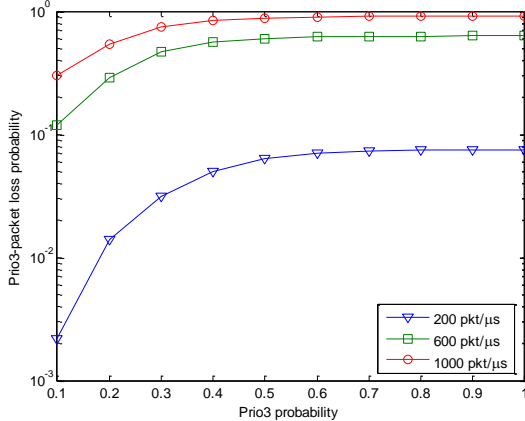
Fig. 13. Highest priority packet loss probability versus probability for a packet to belong to the highest priority class of service.
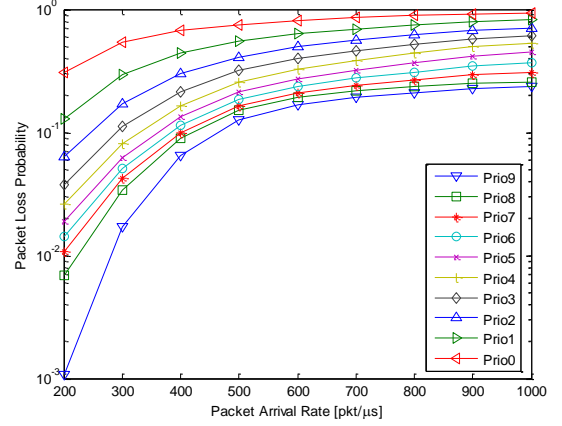
Fig. 14. Packet loss probability versus packet arrival rate for $M = 10$.

Finally, Fig. 14 evaluates the performance of HPF transmission when a higher number of service classes is used. In particular, the figure depicts the case of $M = 10$, with the following packet priority probability distribution function: $p_9 = 0.01$, $p_8 = 0.025$, $p_7 = 0.04$, $p_6 = 0.055$, $p_5 = 0.07$, $p_4 = 0.09$, $p_3 = 0.11$, $p_2 = 0.15$, $p_1 = 0.2$, $p_0 = 0.25$. The capability of HPF transmission in providing service differentiation is confirmed also in this traffic scenario.

V. CONCLUSIONS

This paper proposes the *High-Priority First (HPF) transmission* scheme, a preemption-based service differentiation solution for just-in-time based optical burst switching capable of being both efficient and effective. The proposed method consists in transmitting bursts in chain, in order of decreasing priority. This particular transmission scheme sensibly reduces the preemption probability without affecting the effectiveness of the service differentiation (i.e., a burst is always transmitted when contends with a lower priority one). Analytical and simulation results confirm these properties and also show the effectiveness of the proposed method with a high intensity of high priority traffic and with a large number of service classes. The paper also shows how the proposed solution outperforms the burst cluster transmission scheme, a non-preemptive contention resolution method from which the adopted burst assembly algorithm is derived and which represents the state-of-the-art solution concerning the management of service differentiation in optical burst switching networks.

As said, HPF transmission avoids the adoption of probabilistic preemptive schemes at core nodes (i.e., contention resolution methods in which a high priority burst preempt a lower priority one only with a given probability factor) in order not to affect the effectiveness of service differentiation. However, since such contention resolution schemes can further reduce the overall preemption probability, future work will be devoted to investigate the real implications that these methods could have when used in conjunction with the HPF algorithm. Network performance and effectiveness of service differentiation will be analyzed, and ad-hoc algorithms will be studied and validated to maximize the effects of

service differentiation at given values of the probability factor adopted at core nodes. Future work will also include the definition of proper models to analytically study the behavior of an entire HPF network.

REFERENCES

[1] C. Qiao and M. Yoo, "Optical Burst Switching (OBS) – A new paradigm for an optical Internet," *Journal of High Speed Networks*, vol. 8, no. 1, pp. 69-84, Jan. 1999.

[2] C. Qiao, "Labeled optical burst switching for IP-over-WDM integration," *IEEE Communications Magazine*, vol. 38, no. 9, pp. 104–114, Sept. 2000.

[3] Y. Chen, C. Qiao, and X. Yu, "Optical Burst Switching: a new area in optical networking research," *IEEE Network*, vol. 18, no. 3, pp. 16-23, May 2004.

[4] J.S. Turner, "Terabit burst switching", *Journal of High Speed Networks*, vol. 8, no. 1, pp. 3-16, 1999.

[5] K. Dolzer, C. Gauger, J. Spath, and S. Bodamer, "Evaluation of reservation mechanisms for optical burst switching," *International Journal of Electronics and Communications*, vol. 55, no. 1, pp. 18–26, 2001.

[6] B. Kim, Y. Cho, J. Lee, Y. Choi, and D. Montgomery, "Performance of Optical Burst Switching Techniques in Multi-Hop Networks," in *Proc. IEEE Int. Conf. on Global Communications (GLOBECOM)*, Taipei, Taiwan, pp. 2772-2776, Nov. 2002.

[7] A. Detti, V. Eramo, and M. Listanti, "Performance evaluation of a new technique for IP support in a WDM optical network: optical composite burst switching (OCBS)," *IEEE/OSA Journal of Lightwave Technology*, vol. 20, no. 2, pp. 154-165, Feb. 2002.

[8] A. Sahara, K. Shimano, K. Noguchi, M. Koga and Y. Takigawa, "Demonstration of Optical Burst Data Switching Using Photonic MPLS Routers Operated by GMPLS Signaling," In *Proc. IEEE Int. Conf. on Optical Fiber Communications (OFC)*, Atlanta, GA, Mar. 2003.

[9] R. Ramaswami, K. N. Sivarajan, "Optical Networks: a Practical Perspective," *Morgan Kaufmann*, San Francisco, CA, 2002.

[10] B. Mukherjee, D. Banerjee, S. Ramamurthy, and A. Mukherjee, "Some principles for designing a wide-area WDM optical network," *IEEE/ACM Transactions on Networking*, vol. 4, pp. 684-696, Oct. 1996.

[11] J. Y. Wei, J. L. Pastor, R. S. Ramamurthy, and Y. Tsai, "Just-in-time optical burst switching for multi-wavelength networks," in *Proc. IFIP Int. Conf. on Broadband Communications (BC)*, Hong Kong, China, pp. 339–352, Nov. 1999.

[12] J. Y. Wei and R. I. McFarland, "Just-In-Time signaling for WDM Optical Burst Switching networks," *IEEE/OSA Journal of Lightwave Technology*, vol. 18, no. 12, pp. 2019-2037, Dec. 2000.

[13] I. Baldine, M. Cassada, A. Bragg, G. Karmous-Edwards, D. Stevenson, "Just-in-Time Optical Burst Switching Implementation in the ATDnet All-Optical Networking Testbed," in *Proc. IEEE Int. Conf. on Global Communications (GLOBECOM)*, San Francisco, CA, pp. 2777-2781, Dec. 2003.

[14] M. Yoo, M. Jeong, and C. Qiao, "A high-speed protocol for bursty traffic in optical networks," in *Proc. SPIE Int. Conf. on All-Optical Networking: Architecture, Control and Management Issues*, Boston, MA, pp. 79–90, Nov. 1997.

[15] C. Qiao and M. Yoo, "Choices, features and issues in optical burst switching," *Optical Networks*, vol. 1, no. 2, pp. 36–44, Apr. 2000.

[16] J. Teng, G. N. Rouskas, "A Comparison of the JIT, JET, and Horizon Wavelength Reservation Schemes on A Single OBS Node," in *Proc. IEEE Int. Workshop on Optical Burst Switching (WOBS)*, Dallas, TX, Oct. 2003.

[17] V. M. Vokkarane and J. P. Jue, "Prioritized routing and burst segmentation for QoS in Optical Burst-Switched networks," in *Proc. IEEE Int. Conf. on Optical Fiber Communications (OFC)*, Anaheim, CA, Mar. 2002.

[18] L. Yang, Y. Jiang, and S. Jiang, "A probabilistic preemptive scheme for providing service differentiation in OBS networks," in *Proc. IEEE Int. Conf. on Global Communications (GLOBECOM)*, San Francisco, CA, Dec. 2003.

[19] T. Orawiwattanakul, J. Yusheng, "Analysis of probabilistic preemption-based mechanism for QoS provision in OBS networks," *IEEE Int. Conf. on Information Networking (ICOIN 2009)*, Chiang Mai, Thailand, Jan. 2009.

[20] H. C. Cankaya, S. Charcranoon, and T. S. El-Bawab, "A preemptive scheduling technique for OBS networks with service differentiation," in *Proc. IEEE Int. Conf. on Global Communications (GLOBECOM)*, San Francisco, CA, Dec. 2003.

[21] J. Phuritatkul and Y. Ji, "Resource allocation algorithms for controllable service differentiation in Optical Burst Switching networks," *IEICE Transactions on Communications*, vol. I.E88-B, no.4, pp.1424-1431, 2005.

[22] J. Phuritatkul and Y. Ji, "Buffer and bandwidth allocation algorithms for quality of service provisioning in WDM optical burst switching networks," in *Proc. IEEE Int. Conf. on High Speed Networks and Multimedia Communications (HSNMC)*, Toulouse, France, pp. 912-920, Jul. 2004.

[23] J. Phuritatkul, Y. Ji, and Y. Zhang, "Blocking probability of a preemption-based bandwidth-allocation scheme for service differentiation in OBS networks," *IEEE/OSA Journal of Lightwave technology*, vol. 24, no. 8, pp. 2986-2993, Aug. 2006.

[24] V. M. Vokkerane, Q. Zhang, J. P. Jue, B. Chen, "Generalized burst assembly and scheduling techniques for QoS support in Optical Burst-Switched networks," in *Proc. IEEE Int. Conf. on Global Communications (GLOBECOM)*, Taipei, Taiwan, Nov. 2002..

[25] M. Yoo, C. Qiao, and S. Dixit, "QoS performance of Optical Burst Switching in IP-Over-WDM networks," *IEEE Journal on Selected Areas in Communication*, vol. 18, no. 10, pp. 2062-2071, Oct. 2000.

[26] Y. Chen, M. Hamdi, D. H. K. Tsang, and C. Qiao, "Proportional QoS over OBS networks," in *Proc. IEEE Int. Conf. on Global Communications (GLOBECOM)*, San Antonio, TX, pp. 1510-1514, Nov. 2001.

[27] T. Tachibana, S. Kasahara, "Burst Cluster Transmission: service differentiation mechanism for immediate reservation in Optical Burst Switching networks," *IEEE Communications Magazine*, vol. 44, no. 5, pp. 46-55, May 2006.

[28] G. Marchetto, "Minimizing Preemption Probability to Efficiently Support Service Differentiation in Just-In-Time based OBS Networks," in *Proc. IEEE Int. Conf. on Communications (ICC)*, Beijing, China, pp. 5219-5223, June 2008.

[29] T. Tachibana, M. Ueda, and S. Kasahara, "A preemptive scheme with two-way release message transmission in Optical Burst Switching networks," in *Proc. IEEE Int. Conf. on Global Communications (GLOBECOM)*, Dallas, TX, Nov. 2004.

[30] X. Yu, Y. Chen, and C. Qiao, "A study of traffic statistics of assembled burst traffic in Optical Burst Switched networks," in *Proc. IEEE Int. Conf. on Optical Networking and Communications (OPTICOMM)*, Boston, MA, July 2002.

[31] S. Oh, H. H. Hong, and M. Kang, "A Data Burst Assembly Algorithm in Optical Burst Switching Networks," *ETRI Journal*, vol. 24, no. 4, pp. 311-22, Aug. 2002.

[32] Z. Rosberg, H.L. Vu, M. Zukerman, J. White, "Performance analyses of optical burst switching networks," *IEEE Journal of Selected Areas in Communications*, vol. 21, no. 7, pp. 1187-1197, Sept. 2003.

Guido Marchetto is a post-doctoral fellow at the Department of Control and Computer Engineering of Politecnico di Torino. He received the Ph.D. Degree in Computer and System Engineering from Politecnico di Torino in April 2008.

His research topics are network synchronization, wireless networking, and Quality of Service in packet switched network.