Scheduling algorithms for a slotted packet switch with either fixed or variable length packets[§]

F. Callegati[†], C. Develder[‡], W. Cerroni[†], M. Pickavet[‡],
G. Corazza[†], P. Demeester[‡]

[†]: University of Bologna, Dept. of Electronics, Computer Science and Systems (DEIS)
Viale Risorgimento 2, 40136 Bologna, Italy
tel. no. +39 051 2093089 — fax. no. +39 051 2093053
email: {fcallegati, wcerroni, gcorazza}@deis.unibo.it
[‡]: Ghent University - IMEC, Dept. of Information Technology (INTEC)
Sint-Pietersnieuwstraat 41, 9000 Gent, Belgium
tel. no. +32 9 267 35 93 — fax. no. +32 9 267 35 99
email: {chris.develder, mario.pickavet, piet.demeester}@intec.ugent.be

Abstract: We address the problem of congestion resolution in Optical Packet Switching. We consider a fairly generic all-optical packet switch architecture with a feedback optical buffer constituted of fibre delay lines. Two alternatives of switching granularity are addressed for a switch operating in a slotted transfer mode: switching at the slot level (i.e. fixed length packets of a single slot) or at the burst level (variable length packets that are integer multiples of the slot length). For both cases, we show that in spite of the limited queuing resources, acceptable performance in terms of packet loss can be achieved for reasonable hardware resources with an appropriate design of the time/wavelength scheduling algorithms. Depending on the switching units (slots or bursts), an adapted scheduling algorithm needs to be deployed to exploit the bandwidth and buffer resources most efficiently.

Keywords: IP-over-WDM, Optical Packet Switching, Optical Packet Router, Contention Resolution

Contact author: Prof. Franco Callegati,
DEIS - University of Bologna,
Viale Risorgimento 2,
I-40136 Bologna ITALY
Tel +39 0512093089 Fax +39 0512093053
E-mail fcallegati@deis.unibo.it

[§] This work has been supported by the European Commission through the IST-project DAVID (IST-1999-11387), and by the Flemish Government through the IWT GBOU-project "Optical Networking and Node Architectures". C. Develder is supported as a Research Assistant of the Fund for Scientific Research – Flanders (F.W.O.–Vl.), Belgium

1. Introduction

To date, the presence of optics in telecommunication networks is dominated by point-to-point optical transmission between electronic switching boxes. Over the last couple of years, migration from electrical to optical switching has been commenced in order to match the switching technology to the huge bandwidth capacity offered by (D)WDM transmission. A longer term approach towards real optical networking (as opposed to point-to-point transmission) is Optical Packet Switching (OPS). Compared to circuit switching as applied by wavelength routing approaches, OPS promises to take full advantage of available resources, since it is able to handle traffic at a much finer granularity. Even though today the available bandwidth in backbone networks may seem abundant, the —despite recent economic fallback still steady— growth rate of telco and data traffic will eventually call for more efficient exploitation of the networks' capacity. Indeed a switching granularity at the level of the single wavelength, corresponding to a dedicated bandwidth of up to 40 Gbit/s, may result in a very poor channel utilization that can only be overcome with an efficient traffic grooming, requiring a large number of conversions from and to electrical client layers. OPS alleviates this problem by providing smaller granularity access to the optical bandwidth (on a packet-by-packet basis), thus avoiding the costly electro-optical conversions [1–3]. In addition, the packet switching approach has also advantages with respect to resilience [4], stemming from the easier sharing of resources (i.e., bandwidth) for working and backup purposes.

The actual deployment of OPS is still being hampered by technological difficulties, which mainly comprise lack of optical random access buffers, lack of advanced all-optical packet processing, packet synchronization and optical regeneration. While intermediate solutions between packet and circuit switching, such as Optical Burst Switching (OBS) relax some of the technological requirements, long term approaches should nevertheless envisage packet switching in the optical domain [5]. Currently, the aforementioned difficulties are being addressed in various research laboratories and international projects: all-optical 3R regeneration has been demonstrated (e.g., [6]), all-optical packet header processing (albeit rather basic) has been demonstrated (e.g., [7]), the first approaches to optical synchronization are available (e.g., [8]), and the lack of random access buffers can be strongly alleviated through optical Fibre Delay Lines (FDLs), as illustrated e.g., further in this paper, or through carefully designed Medium Access Control (MAC) protocols with electronic buffers at the network's edges [9]. Given this already promising state-of-the art, we assume that most limitations of optical technology will be overcome.

Moreover, to demonstrate the advantages brought forward by OPS, and to investigate its (technological) feasibility, multiple OPS test networks have been developed and evaluated [10–14]. In this framework DAVID, "Data And Voice Integration over DWDM", is a research project funded by the European Union IST (Information Society Technology) program, aiming at investigating network solutions to offer an optical packet-switched transport. The network proposed by the DAVID project has been conceived to interconnect access points both on a metropolitan and on a backbone scale. Thus, the network architecture is based on a hierarchical structure consisting of several optical Metropolitan Area Networks (MAN) interconnected through an optical Wide Area Network (WAN), all operating in a packet-switched mode.

The WAN consists of a mesh topology of interconnected Optical Packet Routers (OPR). The mesh could rely on a virtual topology based on light-paths in a wavelength-routed network. This overall WAN architecture is well suited to the use of GMPLS for network control and signaling, with a hierarchy ranging from conventional electrical MPLS to optical MPLS and MPλS. This hierarchical structure guarantees scalability and is able to fulfill the needs in terms of level of aggregation and capacity, while offering support for Quality of Service (QoS) at the network level and tools for traffic engineering.

With respect to the packet format, the DAVID consortium has opted for a slotted approach: data bursts from legacy networks are inserted into slots of fixed size that are transmitted in a synchronous way. Two approaches have been studied in the DAVID project: Fixed Length Packet (FLP) and Slotted Variable Length Packet (SVLP). In FLP, each slot has its header and is treated independently from other slots. In SVLP, slots carrying information from the same data burst are kept together, with just one header in the opening slot, and the complete train of slots is treated as a whole.

One of the main engineering issues in the WAN is contention resolution in the OPR. A significant amount of effort has been devoted to this problem and the paper reports a summary of the results obtained by the research groups of INTEC at Ghent University and of DEIS at the University of Bologna. The research performed so far has addressed both the FLP and SVLP approaches, and implications of the choice of the packet format on the congestion resolution strategies have been analyzed. In this paper, we summarize our findings and show how in both cases an appropriate scheduling mechanism can achieve satisfactory performance.

The remainder of the paper is organized as follows. In Section 2, an overview of the OPR architecture is given, followed in Section 3 by an overview of the optical packet formats. In Section 4, a thorough discussion of the problem of contention resolution by the switch's scheduling algorithm is presented and in Section 5 related performance results are reported. Summarizing conclusions are drawn in Section 6.

2. The DAVID Optical Packet Router

The Optical Packet Router proposed within the DAVID project is illustrated in Figure 1. The core of the OPR constitutes a fully non-blocking all-optical switching fabric (based on SOA technology using a broadcast-and-select approach). While the particular technology used for the switching fabric does not impact the scheduling algorithms described further in the paper, an essential feature of the OPR that we will exploit is its wavelength conversion capability, meaning that a signal entering on a particular wavelength can be switched to another wavelength on an output port. The physical implementation of this switching fabric has been recently discussed in [15–17].

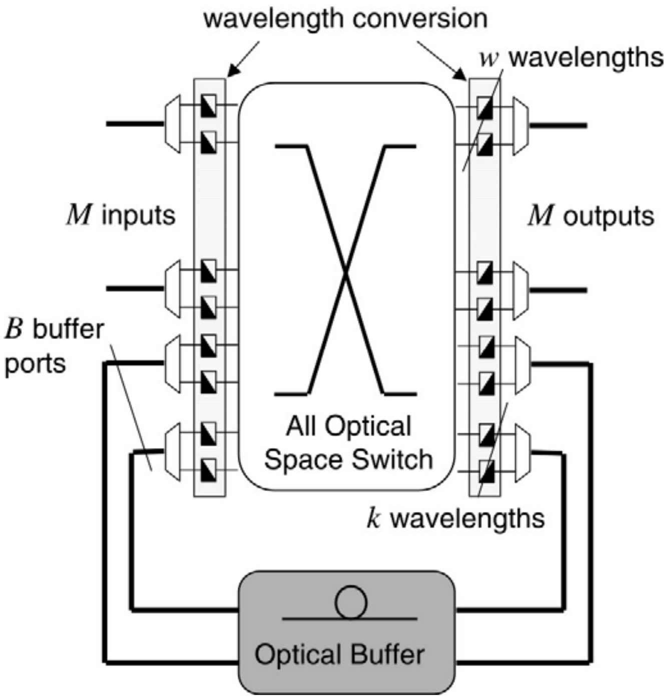

Fig. 1. Schematic of the OPR

The relevant parameters characterizing the OPR are listed in Table I. The inputs and outputs of the switching fabric are connected to *M* input and output fibres, each carrying *w* wavelengths. To help solving contention (see further, Section 4), an optical buffer is implemented using Fibre Delay Lines (FDLs). To allow sharing of these resources, a feedback buffer configuration was chosen: *B*·*k* ports of the switch fabric are used to lead packets to the FDLs, which comprise *B* fibres each carrying *k* wavelengths. After having been delayed in the FDLs, packets will re-enter the switch fabric to have another try at being forwarded to the correct output fibre. The number of times a packet is allowed to be recirculated through these FDLs is denoted by *R*: if the recirculating buffer path does not contain any regeneration of the optical signal, unlimited recirculation would cause irrecoverable signal degradation.

Table I. Relevant parameters of the OPR

Parameter	Meaning
M	number of input and output fibres
w	number of wavelengths per input or output fibre
B	number of FDLs
k	number of wavelengths per FDL
R	number of re-circulations allowed in the FDL buffer
D	granularity of the delay lines, that is the time unit of the delays

For the buffer architecture, we have focused on two possible architectures: F-FDL and I-FDL, as depicted in Figure 2. The first option, F-FDL uses only a single Fixed FDL length for the buffer. Thus, in case of F-FDL, all buffer ports are equivalent, since they all realize the same delay *D*. This implies that to provide longer delays, recirculations will be required. The I-FDL architecture consists of delay lines of increasing lengths, which are all multiples of the basic unit (i.e., delays *D*, *2D*, … *BD*).

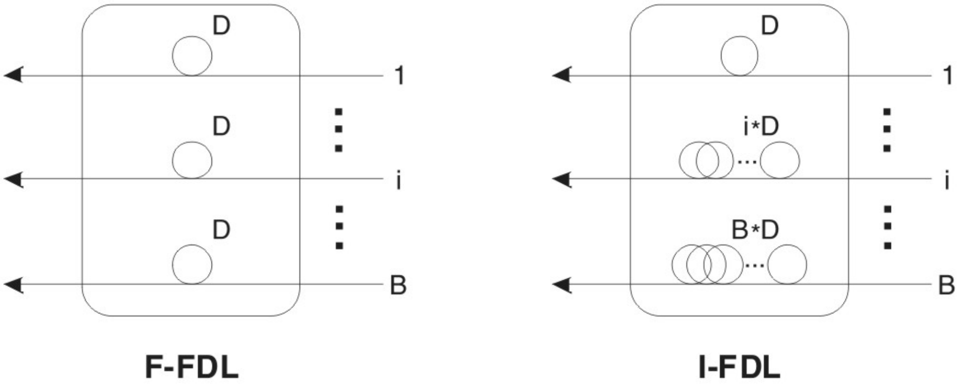

Fig. 2. Buffer architectures: F-FDL versus I-FDL

From a control complexity point of view, the F-FDL buffer seems simpler, since all recirculating ports are equivalent: there is no need to find out what buffer port is most suited to delay a particular packet. However, from a logical performance point of view (e.g., in terms of packet loss), we expect to attain better results with the I-FDL structure, for the same number of recirculation ports $P=B{\cdot}k$. Indeed, the value of *P* indicates how many packets may be entered into the buffer each timeslot. Clearly, when a packet needs to be delayed for a longer time, the F-FDL structure will require that this packet is recirculated through the FDLs multiple times, repeatedly occupying one of the *P* access places to the buffer. In case of I-FDL, the same delay could be realized using only a single passage through the I-FDL buffer. Another way of understanding the better performance of I-FDL is by considering that it can contain more data: the accumulated fibre length in an I-FDL buffer is longer than that of an F-FDL buffer. To obtain the same performance we expect a switch with an F-FDL buffer would require many more recirculation ports B·k, implying, at the least, a larger cost in term of hardware complexity (cf. requiring a larger switch fabric).

Once the architecture has been chosen, an important parameter that needs to be set is the length of the fibre delay lines (in case of F-FDL, all fibres will have length D; in case of I-FDL various delays are available with a delay granularity D). A study of the optimal choice of delay granularity compared to packet length in case of variable length packets for an output-buffered switch has been discussed in [18]. The principal conclusion is that there is an optimal choice for the delay granularity: when it is too small (compared to the average packet length), the chance a packet can be delayed long enough for the contention problem to have disappeared is small. On the other hand, when D is large compared to the packet length, packets will be delayed longer than necessary, creating gaps between successive packets (i.e., periods of inactivity) that limit the utilization of the available bandwidth. We expect to see the same effect in case of SVLP with the feed-back buffer studied here.

With respect to the choice of *B* and *k*, their product will be fixed for a given number of switch fabric ports dedicated to recirculation through the FDL buffer. If $k = 1$, then the I-FDL buffer can realize a broader range of delays, but there is no chance to delay multiple packets of the same amount. By increasing *k*, the range of delays is more limited, but there is the opportunity to delay multiple packets, that are entered simultaneously to the buffer, for the same amount of time. Clearly, when maximizing *k*, we fall back to the F-FDL case, with only a single delay.

3. The DAVID Optical Packet Format

Within the DAVID project, a slotted switching approach was chosen in both MAN and WAN: each timeslot, the incoming slots are switched simultaneously to the outputs. It is indeed true that this choice requires packet synchronization (note that this does not imply bit level synchronization) and alignment, not easy tasks in the optical domain, but has also several advantages since it simplifies the implementation of many classical telecom techniques such as traffic shaping, load balancing, flow control mechanisms, and traffic differentiation. Moreover it is well known that queuing is more effective in this case — an important issue in the optical world with only limited buffer space. Also, at the packet level, a slotted synchronous operation simplifies the implementation of the switching fabric, since it relaxes the need of being fully non-blocking to re-arrangeably non-blocking [19]. Thus, also for the OPR in DAVID a slotted, fixed-length packet approach, has been adopted. It is assumed that the network carries traffic in optical slots of duration *T* and that synchronization units are available at the input interface of the OPR to allow synchronous operation of the switching fabric.

With a slotted packet format the issue arises of fitting the client data bursts in the optical slots. Assuming that *T* is in the order of 1 μs, at a link speed of 10 Gbit/s an optical slot will be able to carry in the order of 1 Kbyte of data. Assuming that the client data packets will be of variable length (mainly IP traffic), ranging from shorter to possibly longer sizes than an optical slot, both the issues of traffic grooming and of segmentation arise. Grooming aims at aggregating short packets on the same data paths, to fill as much as possible the payloads of the optical slots, and segmentation will take care to split long IP packets into several optical slots. The aim of this paper is not to analyze these issues in detail, which constitute a separate set of network engineering problems, that have been analyzed for instance in [20].

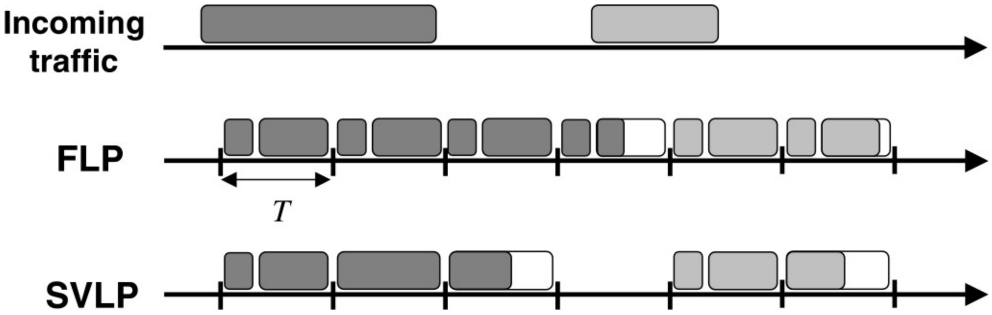

Fig. 3. Comparison between the FLP and SVLP approaches. The lighter shaded areas at the packet's tail represent wasted space stemming from padding

For the discussion at hand, when segmentation is required the alternatives available are to use a Fixed Length Packet (FLP) or a Slotted Variable Length Packet (SVLP) approach, as explained in the introduction. SVLP limits the processing in the optical nodes. It is well known that header processing, including making the forwarding decisions, is a critical issue in an optical packet router because of the very high speed links, implying a high packet arrival rate [21]. With SVLP the routing information is inserted only in the first slot of the train and the whole train is then processed as a whole according to this information, as indicated in Figure 3.

The mean load on the control logic of the switch is thus reduced by a factor that is roughly proportional to the average number of packets per train. Also the overhead due to header information is reduced by a similar factor, since fewer slots carry the control information of the header. FLP on the other hand offers more flexibility in the utilization of the output links and in contention resolution, since forwarding decisions are taken for each individual slot rather than per train. This implies that FLP has the potential of using the available resources more efficiently, since the scheduling algorithm will have more freedom to treat the individual slots. However, with respect to amount of reassembly-work at the edges of the OPS network to the client IP layer, the segmentation and reassembly functions are likely to be more complex (cf. multiple slots composing a single IP packet may have been reordered, in contrast to the SVLP approach that will keep those slots together and in order). In both FLP and SVLP cases, reordering of client layer traffic (e.g., the IP packets) can occur.

4. Scheduling algorithms for recirculating FDL buffer

The task of the OPR is to forward optical packets from its inputs to its outputs. This task is partitioned into (i) *forwarding* and (ii) *switching*. The former implies making the decision where a packet will be forwarded to (i.e. which fibre), based on the information contained in the packet's header, e.g., address (or label) matching with a forwarding table. The latter is the physical action of transferring a packet to the proper output interface. Since optical data processing to date is still very limited, also in DAVID a pragmatic approach to OPS is taken, using a combination of electronics and optics:

- *forwarding* is performed in electronics, converting the header from the optical to the electrical domain;
- *switching* is done all-optically, i.e., without opto-electronic conversion.

Once the header has been read, and it is known what fibre the packet needs to be forwarded to, the OPR's Switch Control Logic (SCL) must decide how to manage the optical payload in order to properly perform the switching. It is at this level that contention has to be faced, which occurs when multiple packets need to be switched to the same output fibre at the same time. In conventional electronic routers, this contention problem is solved through the use of electronic RAM. The lack of optical RAM makes this unfeasible: the only available form of optical "buffers" to date is the use of Fibre Delay Lines (FDLs), where the time a packet is kept in the FDL is predetermined and is directly related to the fibre length. This is in sharp contrast with electrical routers, where a packet is buffered in a queue and retrieved when the output becomes available, without knowing in advance when this will be. The SCL of the optical switch can only let a packet out of the buffer at a particular time: after it has traversed the full FDL length. Still, the limited buffer space and its non-random access nature can be partially overcome by wavelength multiplexing to achieve acceptable Packet Loss Rates (PLR). Thanks to (D)WDM, a single output fibre carries many wavelengths and packets contending for the same output can be multiplexed on different wavelengths and be transferred over the same fibre at the same time. Thus, when wavelength conversion is available, the (D)WDM dimension can be exploited to assist in solving contention problems. The approach adopted in DAVID is to use a combination of wavelength conversion and optical buffering through equipping the OPR with wavelength converters and some re-circulating fibres (recall Figure 1). The resulting contention resolution approach is illustrated in Figure 4.

From this brief discussion it follows that the SCL task comprises two sub-problems:

- *wavelength allocation*: determine what wavelength will be used to transmit a particular packet, which boils down to selecting a wavelength among the *w* available on the output fibre;
- *delay allocation*: if none of the wavelengths is available, a packet will be sent to one of the FDLs for buffering, thus a decision procedure is needed that determines what delay a packet will be given.

In the following we will number the FDLs in increasing order with their length. Therefore d_1 will be the delay realized by the shortest fibre delay line, d_2 the delay of the second shortest etc. up to d_B that will be the delay of the longest delay line. We will also assume that these delays are consecutive multiples of the same unit D, therefore $d_i=iD$ with $i=1..B$.

Before describing the scheduling algorithms used in this paper for fixed, resp. variable length packets in Subsections 4.2 resp. 4.3, we will first give an overview of related work in the area of slotted packet scheduling.

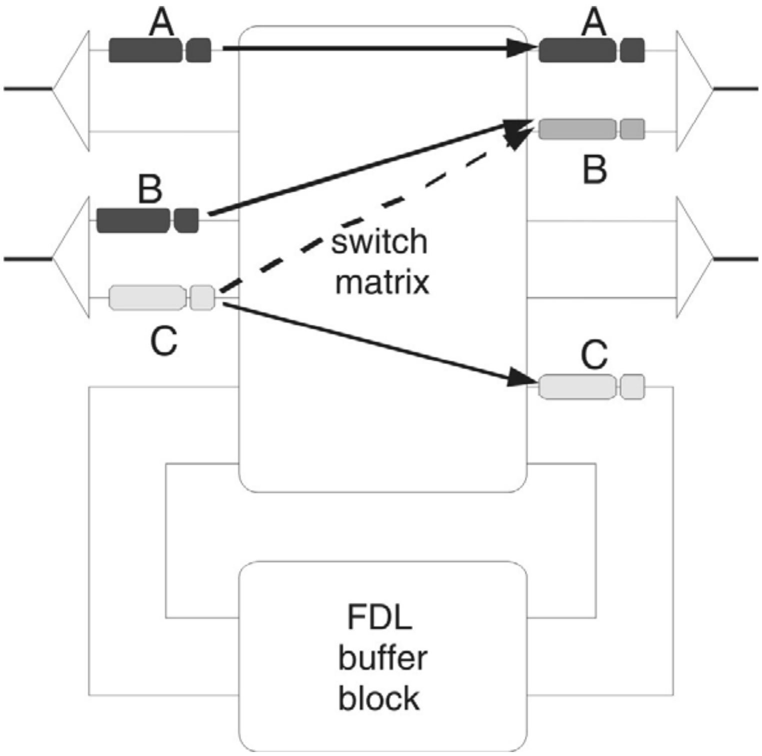

Fig. 4. Illustration of contention resolution approaches in the DAVID network. To accommodate for packet B, the wavelength domain is exploited. Packet C is redirected to the optical buffer to solve the contention.

4.1 Related work

Obviously, scheduling algorithms for slotted packet switching have been studied rather extensively by other authors. In this section, we give an (obviously non-exhaustive) overview of the principal contributions, highlighting also how our original results described next differ from them.

For fixed length packets, related work has been around for some time, where many of it arose in the context of ATM. Examples are [22] and [23], but since those results apply to electrical buffers assuming random access in time, they are not applicable to the OPS scenario considered here. Probably the most relevant work from that era is [24], which considered a shared feed-back buffer delaying packets for a single slot and developed a lower bound on the loss probability. Still, their analysis is inaccurate for practical optical switch sizes (they considered infinite number of switch ports to simplify the analysis). Famous early work focusing on OPS can be found in [25]. However, the architecture and corresponding scheduling algorithm was with feed-forward buffering and without WDM. Work in the frame of the KEOPS project contains similar work, with non-shared output buffering, but with WDM, showing the advantages of exploiting wavelength conversion to solve contention [26]. Other works, e.g., [27], continued to show that also with limited wavelength conversion, significant performance improvements can be achieved. The scheduling algorithms in these approaches basically amount to choosing the smallest available delay solving the contention problem. While they obviously are related to work reported upon here, they do not consider the effect of using a feed-back buffer, and do not consider sharing of buffer resources: traffic for different output ports does not compete for buffer space. Note that such shared recirculating buffer has been analysed in the case of Photonic Slot Routing (PSR) [28], but this is a concept quite different from OPS, since in PSR all wavelengths are switched jointly: they constitute a multi-wavelength slot with a single destination in the PSR network. Recent work which does consider a recirculating buffer, shared among all output ports in an OPS context can be found in e.g., [29] and [30]. However, there the in- and output ports of the switch are single wavelengths and thus the SCL does not have any wavelength assignment component: the combined effect of exploiting the (D)WDM dimension and buffering to solve contention is not addressed. Moreover, the cited references for feed-back buffers do not address the problem of finding suitable scheduling strategies when not all recirculating ports offer the same delay, as with the I-FDL buffer.

Also the scheduling problem for variable length has been addressed in recent research literature. The work of [31] describes the void filling algorithm for fully asynchronous packet switching, showing considerable performance improvement compared to non-void filling approaches. The scheduling algorithm used there was similar to earlier reported results named “horizon scheduling” [32] or LAUC-VF [33]. The main disadvantage of such void filling approaches is that they require maintaining considerable amount of state information on the voids. Therefore, the approaches proposed in this paper are far less complex, yet are proven to be nevertheless quite efficient. Another distinction with cited earlier work is that we focus on fully shared recirculating buffers rather than dedicated output buffers. This scheduling problem in a switch with recirculating FDLs has been addressed before, e.g., in [34], but there the focus lays on a buffer of the F-FDL type with all recirculation ports offering the same delay. Also, [34] focused on a so-called “PostRes” strategy, where a packet entering the buffer will be treated as a newly arriving packet once it leaves the buffer: no wavelength on the outgoing fibre is reserved in advance. Scheduling strategies proposed in this paper are designed for an I-FDL buffer and rather fall in the “PreRes” category, where decisions for delay and wavelength assignment on the output fibre are made jointly.

In the following sections, we detail the scheduling algorithms adopted for the results presented in this paper.

4.2 Fixed length packets: FLP

Since we consider slotted switching where the slot length equals the packet length (including header and guard bands), wavelength allocation can be done on a slot by slot basis: the SCL has only to assure that no more than *w* packets (both newly arrived or coming back from the re-circulating FDLs) are scheduled for the same output fibre. What packet is selected for what wavelength does not matter, since this decision has no repercussions for the following slots: all packets will have ended by then. Note that this implies that the SCL does not need to maintain any state information when an F-FDL is used.

In case of an I-FDL, the only problem that needs to be solved in the FLP case is the delay allocation. Since the SCL deals independently with every slot, it has no knowledge of what new traffic the following slots will bring and thus does not know a-priori which is the best choice of delay to optimize the performance. The only knowledge about "the future" available to the SCL are the packets buffered in the FDLs, (assuming it has enough memory to store the necessary state information to keep track of this). The problem is illustrated in Figure 5.

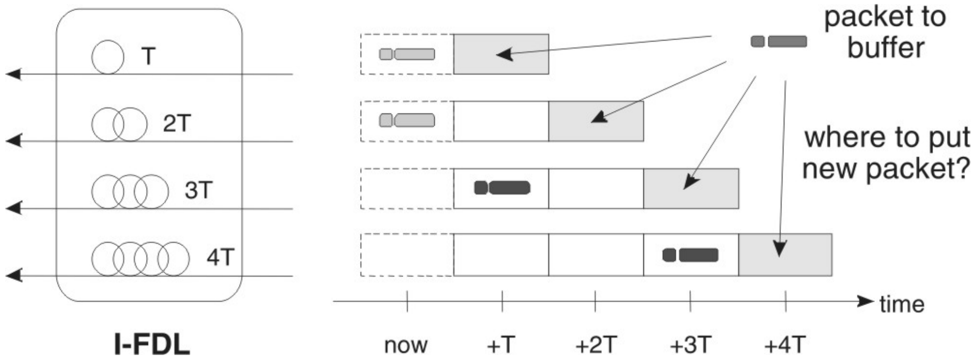

Fig. 5. Illustration of the delay selection strategies for FLP

The following four delay allocation strategies were compared for the FLP case.

- *MinDelay:* for each packet entered in the buffer, the free buffer port with smallest corresponding FDL length is chosen. With respect to Figure 5, this means that we simply start to fill the shaded positions from left to right, regardless of packets that are already in the buffer and will come out at the same time as the newly entered packet. Thus, the sole concern of this strategy is trying to minimize the queuing delay. Note that this was the strategy used by Haas for a feed-forward buffer in [25] and in many works that followed (including [26] and [27]).
- *NoOvr:* the packet is directed to the buffer port leading to FDL *i* with smallest length, such that no more than *w* packets of the same or higher priority will leave the buffer at $now+d_i$ destined for the same output fibre of the OPR; otherwise the packet is dropped. For instance, in Figure 5 we suppose that $w = 1$, and the packets already in the buffer (dark grey) are destined to the same output fibre as the one we have to insert (light grey). The NoOvr strategy implies that we would schedule the new packet for $now+2T$. In the depicted state we would be able to buffer another one for the same output port at $now+4T$, but any more packets for the same output port would be dropped.
- *AvoidOvr:* this strategy combines the two prior ones, by first seeking the free port with smallest FDL length that would not cause overload, but enters the packet at the free port with the smallest delay if no

such overload-avoiding port can be found. Thus, in case of Figure 5 with w = 1, we would first fill *now*+2*T* and *now*+4*T*, and only then *now*+*T* and *now*+3*T*.

- *Balance:* contending packets will be spread in time. To buffer a packet *p*, for each available FDL *i*, the packets scheduled at *now*+d_i directed to the same output port destination as *p* are counted: N_L. The packet is then put in the free FDL with the smallest count N_L. This strategy thus attempts to spread the buffered packets destined for the same output port equally over all timeslots.

Note that for $B \cdot k \leq w$ the first three strategies (all except Balance) amount to the same, since then the packets coming from the buffer structure alone (i.e. without newly arriving packets at the switch's inputs) never can overload an output port.

4.3 Variable length packets: SVLP

For the case of SVLP, the wavelength selection and delay allocation problems are tightly coupled and an algorithm considering the full picture of packets being transmitted and or delayed should be used to obtain an optimal allocation of resources. Unfortunately the design of such an optimal algorithm is very complex, and in contrast with the stringent requirements placed on header processing speed by OPS. Therefore heuristic solutions are used, as discussed in [31] for the case of purely variable length packets.

The main difficulty in solving the wavelength and delay allocation is that due to the variability of the packet length, the voids between two successive packets may not always be large enough to allow a buffered packet to be inserted. Also, due to the discrete delays available from the FDL buffer, buffered packets can not always be inserted directly after the previous packet on the same outgoing fibre, and thus gaps will be created. The problem is illustrated in Figure 6 for the case of four FDLs with four wavelengths each. The shaded packets for each of the wavelengths indicate packets that have been scheduled in the past and are currently under transmission, or have been entered in an FDL and will come out at some point in future (e.g., the packet at t_0+2D on λ_3). Note that these packets are the only ones the SCL can take into account: it has no knowledge of packets that will arrive at the OPR's input ports in future. Thus, with the knowledge depicted in Figure 6, when a new packet arrives at time t_0 and finds all wavelengths busy, the SCL has to decide whether it will delayed through delay line 1 to time t_0+D, or through delay line 2 to time t_0+2D, etc. up to t_0+4D.

A wavelength assignment algorithm that very efficiently tackles this problem is the void filling algorithm presented in [31], although it is very demanding from the computational point of view. According to its name, void filling aims at filling all the gaps in the delay line buffer. Alternative algorithms have been proposed in [31]. Among them are the so-called MINimum Length (MINL) and MINimum Gap (MING) algorithms. They implement a logical FIFO output queuing, and are much simpler than void filling from the implementation point of view. Instead of requiring extensive search in the "list" of existing gaps, which may be fairly large, they just require a direct comparison with the times the various wavelengths on the output fibre will be free. The MINL algorithm chooses the wavelength with the shortest queue while the MING looks for the wavelength where queuing the packet will cause the smallest gap.

In the example of Figure 6, the MINL algorithm will delay the packet of 2D and send it to λ_2, which is the first wavelength to become available for transmission. The MING algorithm will delay the packet of 3D and send it to λ_3 where the gap between the previous and the present packet is minimized. From this example, it is clear that wavelength and delay assignment problems are tackled jointly in case of SVLP: when selecting the wavelength, we also choose when the packet will be put on this output fibre. Once a wavelength has been chosen, the packet is sent to the delay line offering the appropriate delay if this FDL is available, otherwise the packet is dropped.

Although these scheduling algorithms are not new in se, the application of these strategies using a switch with fully shared recirculating buffer, and the investigation of the performance under various parameter settings associated with this buffer structure, are this paper's original contributions.

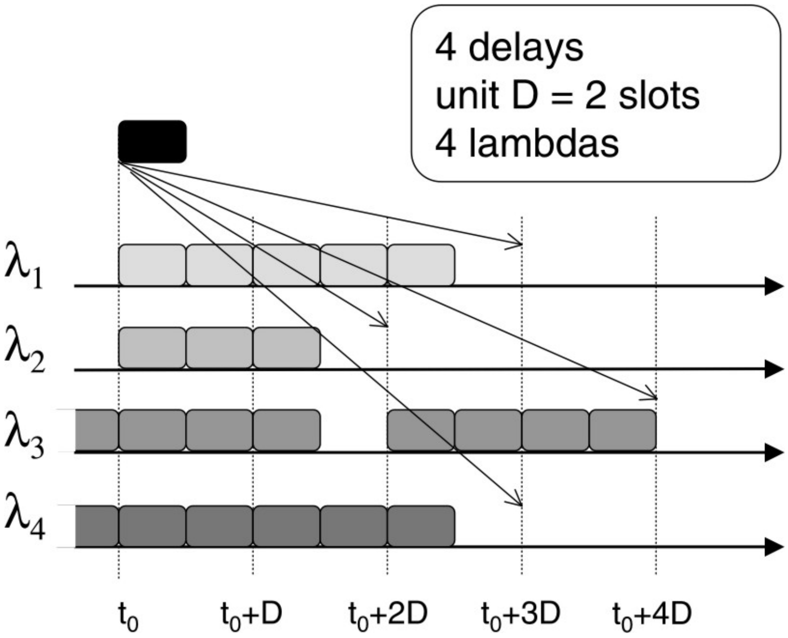

Fig. 6. Illustration of the delay selection strategies for SVLP.

5. Results

To evaluate the different solutions to the wavelength and delay allocation algorithms described before, we have performed a wide range of simulations covering both the FLP and SVLP cases. The results presented in this section were carried out for a common reference configuration of the OPR parameters. As the aim of our studies was to investigate the performance of the algorithms for contention resolution, we varied the parameters for the related resources: the number of recirculating fibres (*B*), and the number of wavelengths on each of them (*k*). The other parameters were fixed, with values in accordance with the range of feasibility of current state of the art technology: we focused on a reference configuration with 6 input/output fibres (M=6), and 32 wavelengths per fibre (*w*=32).

The performance analysis is focused on the Packet Loss Rate (PLR), which is the critical performance parameter in this environment. The very limited queuing capabilities of the delay lines result in very small queuing delays (especially with respect to delays in other parts of the network, e.g., access and metro, or even propagation delays), while dropping of packets is more likely. In the subsequent sections, we will start by motivating the used approach to performance assessment in 5.1, and continue with a discussion of the results for both fixed (Subsection 5.2) and variable (Subsection 5.3) length packets.

5.1 Traffic models and performance assessment methodology

The traffic relevant to the DAVID project is multiplexed from legacy networks by routers and/or switches that interface with the backbone for high speed and long distance site to site interconnection. It is unlikely that the single application or the single user terminal will have direct access to the network. As a consequence traffic analysis and modeling should be kept as general as possible, by means of a few, general theoretical traffic models. We have opted for multiple well-known traffic sources such as the classical Poisson traffic, but to accommodate for more bursty traffic, we have included results for on-off sources with geometrical distributions for the on- and off periods as well. In case of FLP, we also illustrate the performance under self-similar traffic generated through on-off sources with Pareto distributed on- and off periods.

Note that those traffic models can be considered realistic scenarios for the case of an optical packet switched node in the core network: this will be highly multiplexed traffic, collecting many client traffic streams which will be shaped by the optical packet assembly procedures at the network's edges. It is well-known that those packet assembly operations result in e.g., considerable reduction of possible self-similarity present in ingress traffic entering the OPS network [20]. Therefore, we preferred using somewhat classical traffic models rather than tedious modeling of all real-life traffic characteristics, such as TCP closed-loop effects and in- and egress packet (re)assembly. For the purpose of understanding the scheduling algorithm's performance, and various buffer parameters we deemed this to be sufficient.

To assess the performance of the scheduling algorithms, we focused on the operation of a single switch. Analysis and thorough understanding of the various buffer parameters' influence on performance would be hampered by considering network-level scenarios. Since analytical modeling of the switch with recirculating buffer tends to be limited to fairly small switches (e.g., results in [29] and [30]), or either give only lower bounds (as e.g., [24]), and not applicable for all traffic models we considered, we resorted to simulations to assess performance. Obviously, sufficient care has been taken to produce trustworthy results [35] by using reliable random generators and ensuring small confidence intervals (note however that to keep figures sufficiently clear, we have omitted error bars from the graphs).

5.2 FLP

For the performance assessment of the OPR using the FLP approach, we used three different traffic types. The first was the classical Poisson arrival pattern, where the number of simultaneously arriving slots in successive slots is uncorrelated. To investigate the behavior under traffic exhibiting bursty periods of traffic intensity, we used an on/off model where packets arrive back-to-back during the on-periods, and no packets arrive during the off-period. The so-called GeoOnOff model uses the well-known geometric distribution for the duration of both on- and off-periods. The third model, denoted ParetoOnOff, used the heavy-tailed Pareto distribution to generate the period lengths. This particular model was chosen because it is well-known that the aggregate of such traffic sources results in traffic exhibiting self-similarity [37].

For the FLP case the main focus is on understanding which is the most efficient buffer configuration and the best delay allocation strategy. Therefore the simulations have been performed with $D=T$ and allowing an unlimited number of re-circulation.

The first task we set ourselves was to quantify the advantage of using an I-FDL buffer versus F-FDL. The results of this comparison are illustrated in Figure 7 for a load of 0.95. The parameter that we varied was the number of recirculation ports $B{\cdot}k$. For the F-FDL, we use only a single FDL length corresponding to one timeslot (D=1T) for each of the k wavelengths (B=1). For the I-FDL, each of the B FDLs carries only a single wavelength (k=1). As expected, the use of an I-FDL buffer achieves a lower PLR (since the difference in "storage" capacity increases for increasing number of recirculating buffer ports).

With respect to the different traffic types, as in classical queuing, the effect of adding buffer space (in this case adding more recirculating buffer ports) is more effective for traffic showing no correlation: the PLR is considerably lower for Poisson traffic. Still, also for relatively short-time correlations, performance improvement is still considerable (the average on-period in the GeoOnOff case was set to 4 slots). For the self-similar ParetoOnOff model, exhibiting long-range correlations, it is clear that adding a buffer ranging over a small timescale compared to the timescale of correlation cannot achieve much performance improvement. Fortunately, the expectation of encountering self-similar traffic in the DAVID core network is low: the WAN traffic is highly aggregated traffic, and the aggregation process has been shown to eliminate the long-range correlations [20].

The key question we addressed with our FLP simulations was of course which of the proposed strategies (MinDelay, NoOvr, AvoidOvr, Balance; see Section 4) achieved the best results in terms of packet loss. The result of the comparison is depicted in Figure 8. Again, the parameter we varied was the number of recirculation ports. The buffer we chose was the I-FDL as indicated before (k = 1, varying B, with respective FDL lengths L = T,2T…BT). Note that for $B{\cdot}k{\leq}w$=32, the first three strategies (MinDelay, NoOvr and AvoidOvr) are identical, since only for larger number of recirculation ports the "overload" situation can occur. Yet, also for the larger B values, the differences in terms of achieved PLR are minimal. The only strategy that achieves improvement over the simple MinDelay strategy is Balance. The reason for the better performance is obvious: by spreading the packets destined for the same output fibre equally over time, the number of times a packet needs to be recirculated is limited.

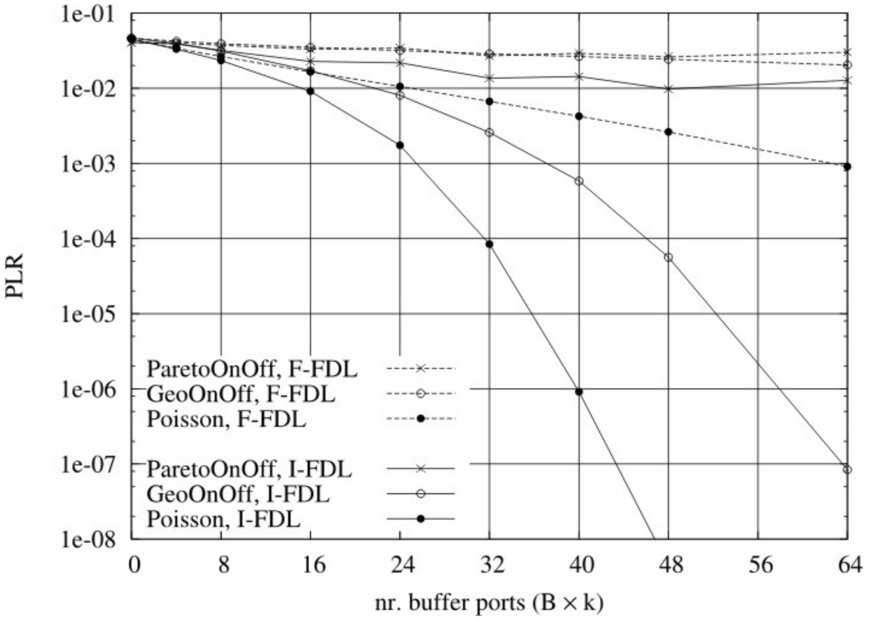

Fig. 7. Comparison of I-FDL (Full lines) and F-FDL (dashed lines) for the case of FLP. The load was set to 0.95.

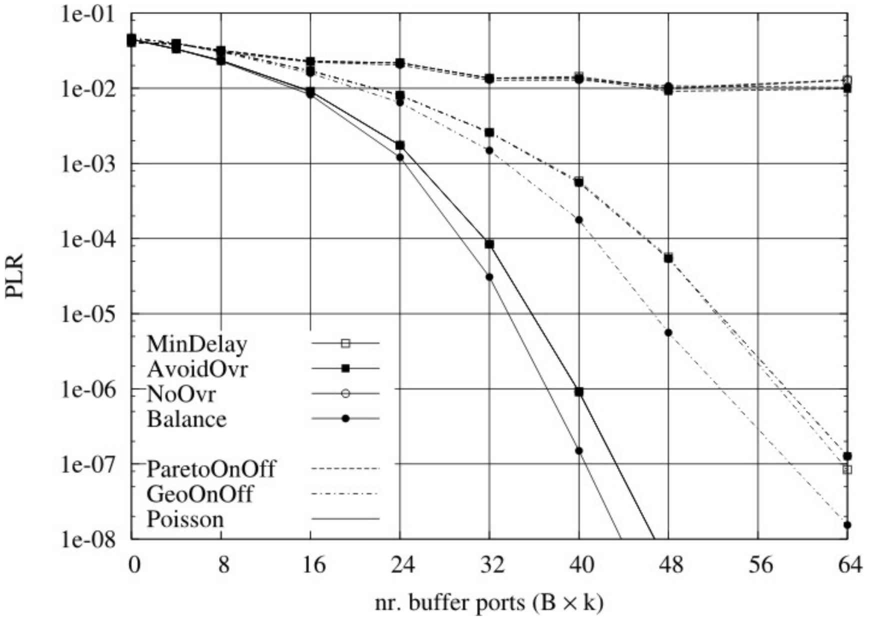

Fig. 8. Comparison of the delay allocation strategies for FLP in case of I-FDL buffer. The load was set to 0.95.

5.3 SVLP

To analyse the performance of the OPR with SVLP, the input traffic has been generated as the output of an M/M/1 queue. Packets arrivals are generated according to a Poisson point process and the packet lengths are exponentially distributed. These packets are fed into a queue; according to Burke’s theorem the resulting output traffic is still a Poisson point process for what regards the arrivals but the packet are “serialized” as it should be in serial transmission line. This traffic is then synchronized to the network reference time slot, padded accordingly and taken as the input for one of the input wavelengths. The average length of the packets is therefore geometrically distributed with an average of *L*=5 slots.

The results here presented focus on the case *B*=2, with the two delay lines taken of size *D* and 2*D*, where *D* is a parameter to dimension. For the time being we also assume that no recirculation is allowed (*R*=1).

As already explained in this scenario the choice of wavelength and delay are related and a separate analysis of the role of the single parameters is almost impossible. The dimensioning of *B*, *k*, and *D* has to be considered altogether. We have chosen, among the possible alternatives, to provide curves that plot the PLR as a function of *k*. In Figure 9 the PLR is plotted as a function of *k*, varying D and comparing the MING and MINL algorithm. The curves in the figure exhibit the same shape, with the PLR that steadily decreases with *k*, up to a point where it flattens up and remains almost constant. This behavior is independent of the algorithm chosen, of the size of the buffer or of the values of the buffer granularity, and is just a function of *k*. It can be explained considering that the loss of packets is due to two concurrent phenomena, the lack of delays and the lack of wavelengths to accommodate a packet at a given delay. The lack of wavelengths is the leading loss phenomena when *k* is small; when increasing *k*, the loss because of lack of delay emerges and it becomes the only responsible of loss when *k* is large enough. Therefore the knee of the curves in this figure provides an immediate estimate of the amount of wavelengths that are needed in the re-circulating fibres in order to fully exploit their delay for congestion resolution. Both graphs present several curves varying the delay granularity *D* as a parameter. It is evident that the larger *D* the better, with an improvement of several order of magnitudes. The price to pay for the improvement in performance is the shift of the knee of the curves towards larger values of *k*, meaning that more wavelengths are needed to exploit the possibilities of congestion resolution offered by larger delays.

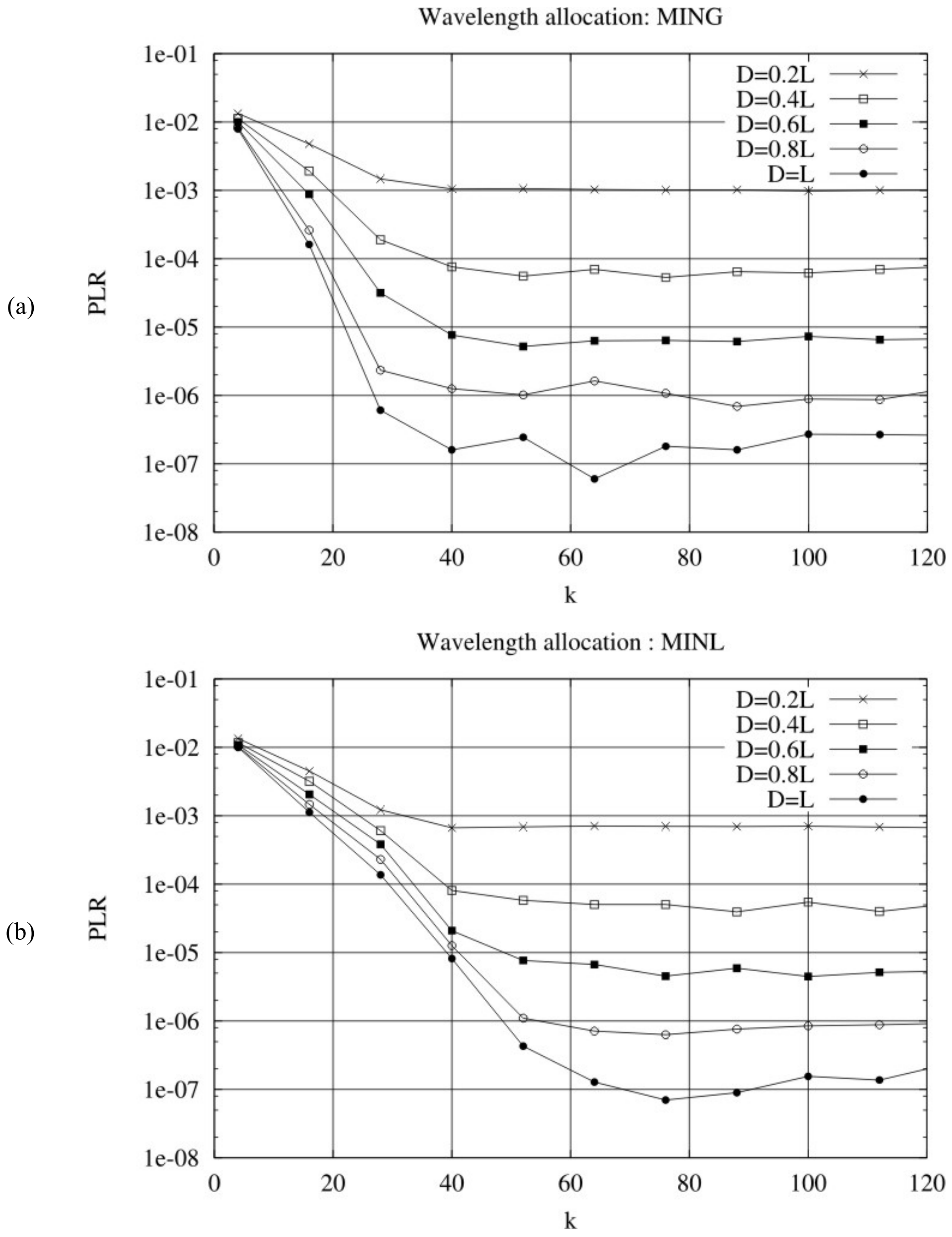

Fig. 9. PLR as a function of k for the MING, resp. MINL, algorithm, L=5T, B=2, M=6 and w=32 and load equal to 0.8.

It is interesting to note that the MINL and MING algorithms perform similarly, at least in this case with a limited number of delays. Nevertheless the MINL requires more wavelengths in the buffer. To reach a PLR of about 10^{-7}, that is the performance for the case with $D=L$, almost twice as many wavelengths are needed in the MINL case than in the MING one.

These two graphs support answers to most of the questions regarding dimensioning the OPR for the SVLP case. We can say that when the number of wavelengths per link fibre is large, the number of re-circulating fibre does not need to be very large. With the parameters used here an I-FDL buffer with just two delay lines seems to be enough. Nonetheless to fully exploit these delay lines, it is very important to supply a sufficient number of wavelengths (buffer ports) and properly dimension the delays. In particular, the unit of delay *D* has to be in the same range as the average train size. The amount of wavelengths *k* is significantly smaller with the MING than with the MINL wavelength allocation algorithm. We can therefore conclude that the MING algorithm is preferable and that with *D*=*L* and *k*=32 the packet loss probability can be kept into a range of fairly acceptable values.

If recirculation of packets is allowed we expect a further improvement in the performance, because, as a matter of fact, more delays will be available. For instance with *B*=2 a packet may be given a delay of 6*D* by making it travel through fibre 2 with delay d_2=2*D* for three times. The price to pay will be in terms of wavelengths since more packets will be traveling the buffer at the same time (in the former example the packet considered will cross the buffer 3 times). In Figure 10, support for this statement is provided. The curves are provided as a function of *k* for the MING algorithm and for the reference configuration M=6, w=32. A F-FDL buffer (B=1) is assumed to de-couple the effect of re-circulation from the effects due to the availability of several delays, as was done in the previous section about FLP. As expected the PLR improves with increasing *R* at the price of more wavelengths in the buffer. Again the PLR flattens up for a large value of k, for the reasons already discussed previously. Considering that the total number of buffer ports is given by the *Bk* product, comparing Figures 9 and 10 clearly show that the simpler F-FDL buffer with re-circulations can be traded for the more complex I-FDL buffer only at the price of a larger number of buffer ports. For instance a value of *Bk*≅120 buffer ports are required to achieve optimal performance with *B*=1 (F-FDL) and *R*=3 while just *Bk*≅80 buffer ports are required for *B*=2 (I-FDL). This statement is in favor of a I-FDL buffer configuration with a few delays also for the SVLP case here analyzed.

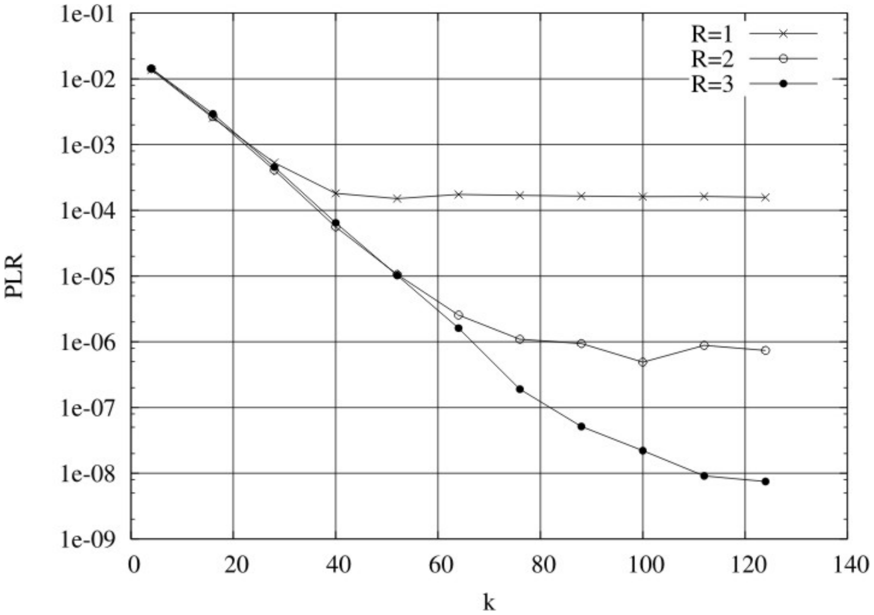

Fig. 10. PLR as a function of k for the MING algorithm, D=L, B=1, M=4 and w=16 and load equal to 0.8.

6. Conclusion

In this paper, we have studied the problem of congestion resolution in an optical packet switch dealing with synchronous and fixed size packets. Congestion resolution is performed by means of wavelength multiplexing thanks to DWDM and by some re-circulating ports equipped with delay lines.

Both the case where the resource allocation and congestion resolution is performed at the single slot level, and the case where slots belonging to the same, segmented, client data burst are kept together and switched as a whole have been considered.

Scheduling algorithms and related numerical results have been discussed, providing guidelines to dimension the critical system parameters in both cases.

The main conclusion of this work is that acceptable performance in terms of packet loss rate can be achieved at the price of a reasonable amount of hardware resources, as long as the proper scheduling algorithms are used to exploit these resources at their best.

References

[1] C. Guillemot et al., "Transparent Optical Packet Switching: The European ACTS KEOPS Project Approach", J. of Lightwave Techn., Vol. 16, No. 12, December 1998, pp. 2117-2134.

[2] D.K. Hunter, I. Andonovic, "Approaches to Optical Internet Packet Switching", IEEE Comm. Magazine, Vol. 38, No. 9, September 2000, pp. 116-122.

[3] S. Yao, B. Mukherjee, S. Dixit, "Advances in Photonic Packet-Switching: An Overview", IEEE Comm. Magazine, Vol. 38, No. 2, February 2000, pp. 84-93.

[4] D. Colle, A. Groebbens, P. Van Heuven, S. De Maesschalck, M. Pickavet, P. Demeester, "Porting MPLS-recovery techniques to the MPλS paradigm", Special Issue on 'Protection and Survivability' of Optical Networks Magazine, Vol. 2, No. 4, July/August 2001, pp. 29-47.

[5] A. Bianco, E. Leonardi, M. Munafò, F. Neri and W. Picco, "Design of optical packet switching networks", Proc. IEEE Global Telecommunications Conference (Globecom 2002), Taipei, Taiwan, November 2002, Vol. 3, pp. 2752-2756.

[6] B. Dagens, A. Labrousse, R. Brenot, B. Lavigne, M. Renaud, "SOA-Based Devices for All-Optical Signal Processing", Proc. Conf. on Optical Fibre Communications (OFC 2003), Atlanta, GA, March 2003, Vol. 2, pp. 582–583.

[7] A. Srivatsa, H. de Waardt, M. T. Hill, G. D. Khoe, and H. J. S. Dorren, "All-optical serial header processing based on two-pulse correlation", Electron. Lett., Vol. 37, No. 4, 15 February 2001, pp. 234–235.

[8] N. Le Sauze, D. Chiaroni, O. Rofidal, A. Dupas, "New optical packet synchronizer for optical packet routers", Proc. Photonics in Switching (PIS 2001), Monterey, CA, June 2001.

[9] A. Bianco, G. Galante, E. Leonardi, F. Neri, and M. Rundo, "Access Control Protocols for Interconnected WDM Rings in the DAVID Metro Network", Proc. Tyrrhenian Intern. Workshop on Digital Communications (IWDC 2001), Taormina, Italy, September 2001, pp. 38-55.

[10] F. Masetti et al., "High Speed, High Capacity ATM Optical Switches for Future Telecommunication Transport Networks", IEEE J. Selected Areas in Commun., Vol. 14, No. 5, June 1996, pp. 979-998.

[11] P. Gambini, "Transparent optical packet switching: network architecture and demonstrators in the KEOPS project", Invited paper, IEEE Journal of Selected Areas in Communications, Vol. 16, No. 7, September 1998, pp. 1245-1259.

[12] D. Wonglumson, I. M. White, S. M. Gemelos, K. Shirkande, L. G. Kazovsky, "HORNET- A Packet-Switched WDM Network: Optical Packet Transmission and Recovery", IEEE Phot. Techn. Lett., Vol. 11, No. 12, December 1999, pp. 1692-1694.

[13] D. Hunter, M. Nizam, M. Chia, I. Andanovic, K. Guild, A. Tzanakaki, M. O'Mahony, J. Bainbridge, M. Stephens, R. Penty, I. White, "WASPNET: A Wavelength Switched Packet Network ", IEEE Comm. Magazine, Vol. 37, No. 3, March 1999, pp. 120-129.

[14] I. Chlamtac, et. al., "CORD: Contention Resolution by Delay Lines", IEEE J. Selected Areas in Commun., Vol. 14, No. 5, June 1996, pp. 1014-1029.

[15] D. Chiaroni, et al., "First demonstration of an asynchronous optical packet switching matrix prototype for MultiTerabit-class routers/switches", in Proceedings of 27th European Conference on Optical Communication (ECOC 2001), Amsterdam, The Netherlands, October 2001, Vol. 6, pp. 60-61.

[16] N. Sahri, et al., "A highly integrated 32-SOA gates optoelectronic module suitable for IP multi-terabit optical packet routers", Post-deadline paper, OFC 2001, Anaheim, CA, March 2001.

[17] F. Masetti et al, "Design and Implementation of a Multi-Terabit Optical Burst/Packet Router prototype", Postdeadline paper, OFC 2002, Anaheim, CA, March 2002.

[18] F. Callegati, "Optical Buffers for Variable Length Packets", IEEE Communications Letters, Vol. 4, No.9, September 2000, pp. 292-294.

[19] C. Develder, J. Cheyns, E. Van Breusegem, E. Baert, A. Ackaert, M. Pickavet, and P. Demeester, "Node architectures for optical packet and burst switching", Proc. Int. Topical Meeting on Photonics in Switching (PS2002), (invited) paper PS.WeA1, Cheju Island, Korea, 21–25 Jul. 2002, pp. 104–106.

[20] H. Christiansen, Using OPNET to compare and analyse different traffic-bundling schemes, Proceedings of OPNETWORK 2001, Washington, D. C., USA, August 2001.

[21] F. Callegati, H. C. Cankaya, Y. Xiong, M. Vandenhoute, "Desing issues for optical IP routers", IEEE Comm. Magazine, Vol. 37, No. 12, December 1999, pp. 124-128.

[22] G. Bianchi, J.S. Turner, "Improved queueing analysis of shared buffer switching networks", IEEE/ACM Transactions on Networking, Vol. 1, No. 4, August 1993, pp. 482-490.

[23] A. Monterosso, A. Pattavina, "Performance Analysis of Multistage Interconnection Networks with Shared-Buffered Switching Elements for ATM Switching", Proc. IEEE Infocom 1992, Florence, Italy, May 1992, Vol. 1, pp. 124-131.

[24] M.G. Hluchyj, M.J. Karol, "Queueing in High-Performance Packet Switching", IEEE J. on Selected Areas in Commun., Vol. 6, No. 9, December 1988, pp. 1587-1597.

[25] Z. Haas, "The 'staggering switch': an electronically controlled optical packet switch", J. Lightwave Technology, Vol. 5, No. 5, May-June 1993, pp. 925-936.

[26] S.L. Danielsen, C. Joergensen, B. Mikkelsen, K.E. Stubkjaer, "Analysis of a WDM packet switch with improved performance under bursty traffic conditions due to tuneable wavelength converters", J. Lightwave Technology, Vol. 16, No. 5, May 1998, pp. 729-735.

[27] G. Shen, S.K. Bose, T. Hiang Cheng, C. Lu and T. Yoong Chai, "Performance study on a WDM packet switch with limited-range wavelength converters", IEEE Commun. Letters, Vol. 5, No. 10, October 2001, pp. 432-434.

[28] H. Zang, J.P. Jue and B. Mukherjee, "Photonic slot routing in all-optical WDM mesh networks", Proc. IEEE Global Telecommun. Conf. (Globecom 1999), Rio de Janeiro, Brazil, December 1999, Vol. 2, pp. 1449-1453.

[29] A. Kushwaha, Sanjay K. Bose and Y. N. Singh, "Analytical modeling for performance studies of an FLBM-based all-optical packet switch", IEEE Commun. Letters, Vol. 5, No. 5, May 2001, pp. 227-229.

[30] P.D. Bergstrom Jr., M.A. Ingram, A. J. Vernon, J.L.A. Hughes, P. Tetali, "A Markov Chain Model for an Optical Shared-Memory Packet Switch", IEEE Trans. on Commun., Vol. 47 , No. 10, October 1999, pp. 1593-1603.

[31] L. Tancevski, S. Yegnanarayanan, G. Castanon, L. Tamil, F. Masetti, and T. McDermott, "Optical routing of asynchronous, variable length packets", IEEE Journal on Selected Areas in Communications, Vol. 18, No. 10, October 2000, pp. 2084-2093.

[32] J.S. Turner, "Terabit burst switching", J. High Speed Networks, Vol. 8, No. 1, January 1999, pp. 3-16.

[33] Yijun Xiong, Marc Vandenhoute and Hakki C. Cankaya, “Control architecture in optical burst-switched WDM networks”, IEEE J. on Selected Areas in Commun., Vol. 18, No. 10, October 2000, pp. 1838-1851.

[34] C. Gauger, “Dimensioning of FDL Buffers for Optical Burst Switching Nodes”, Proc. Conference on Optical Network Design and Modeling (ONDM 2002), Torino, Italy, February 2002, pp. 117-132.

[35] K. Pawlikowski, H.-D.J. Jeong, and J.-S.R. Lee, “On credibility of simulation studies of telecommunication networks”, IEEE Commun. Magazine, Vol. 40, No. 1, Jan. 2002, pp. 132-139.

[36] F. Callegati, W. Cerroni, G. Corazza “Optimization of Wavelength Allocation in WDM Optical Buffers”, Optical Networks Magazine, Vol. 2, No. 6, November 2001, pp. 66-72.

[37] W. Willinger, M. Taqqu, R. Sherman, and D. Wilson, “Self-similarity through high-variability: statistical analysis of Ethernet LAN traffic at the source level”, IEEE/ACM Trans. Networking, Vol. 5, No. 1, January 1997, pp. 71-86.