Optical slot size dimensioning in IP/MPLS over OPS networks

Davide Careglio, Josep Solé Pareta, Salvatore Spadaro
Universitat Politècnica de Catalunya
Advanced Broadband Communications Labs
Barcelona, Catalunya, Spain, 08034
Email: {careglio, pareta}@ac.upc.es, spadaro@tsc.upc.es

***Abstract*— Current optical technology allows an easy implementation of synchronous, time-slotted optical networks. In contrast, traffic in electrical domain (for instance IP) mainly consists of asynchronous, variable length packets. Therefore a packet format adaptation process is needed between electrical and optical domains, the optical slot size being in this process a crucial parameter, which strongly influences the bandwidth utilisation and determines the overall network performance. In this paper we address the problem of designing this adaptation layer and of dimensioning the optical slot size in IP/MPLS over optical packet switched networks. Three different optical packet formats, namely Fixed-Length Packet, Slotted Variable-Length Packet, and Fixed-Length Packet with Traffic Aggregation are discussed. To find the optimum size, the efficiency of such formats are evaluated by simulations.**

I. Introduction

In the future telecommunication networks is expected that optical technologies and IP will be the dominant solutions for the next generation Internet, reaching world-wide diffusion and acceptance [1]. Thus, packet-based technologies have become a very important area of study, encompassing a wide range of solutions to supporting predominantly IP traffic over WDM optical links. In a long-term scenario, the optical packet switching (OPS) can provide a simple transport platform based on a direct IP over WDM structure which can offer high bandwidth efficiency, flexibility, and fine granularity [2].

In this context, IST project DAVID (IST-1999 11742) aims at proposing a viable approach towards OPS, by developing networking concepts and technologies for future optical networks. The DAVID network encompasses both regional, metropolitan area networks (MAN) and backbone wide-area network (WAN) as shown in Fig. 1 [3]. In both domains, fixed-length packets are used in a synchronous, slotted mode of operation. This is the most studied case in literature and, at the present level of technology, the easiest implementable solution. This because the additional hardware complexity, due to synchronisation units necessary in the nodes, is counterbalanced by a simpler control of the access protocols and switching nodes (e.g. no packet length detection, easier buffering management, etc.) which can provide better performance than asynchronous, variable case [4] [5]. In contrast, traffic in electrical domain (for instance IP) mainly consists of asynchronous, variable length packets [6] which creates a clear format incompatibility with the considered OPS networks.

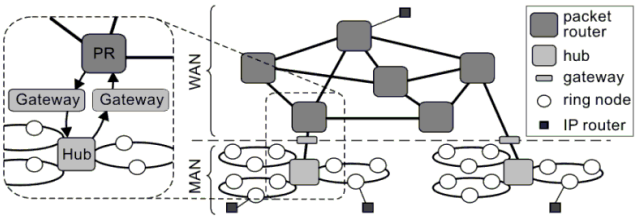

Fig. 1. The DAVID network architecture

In fact, the fitting of variable length data packets into the slots necessarily leads to round off the data packets to an integer number of slots (by means of some padding). As a consequence, the effective amount of information to be carried by the network is greater than the real amount of information included into the data packets. At the same time, the incoming packets arrive at the edge of the optical network in a asynchronous mode, then they are fitted into discrete sized slots and transmitted synchronously to the network. Therefore also the packets inter-arrival time is made discrete by an integer number of time slots which clearly influences the traffic load. Therefore, the design of the DAVID networks (both MAN and WAN parts) needs to address the problem of adapting the packet format between electrical and optical domains at the edge of the optical network.

The purpose of this work is to design this adaptation layer focussing on dimensioning the size of the optical slot, as the crucial parameter that strongly influences the bandwidth utilisation and, hence, the overall network performance. There are really few investigations in literature dealing with this problem. The most accurate study is presented in [7] where the issue of the optimal size of the time slot is addressed, with reference to access delay and traffic shaping. However, it is focused only on an OPS network carrying ATM cells. Here, we address the same issue but considering a currently more realistic IP over OPS scenario and we focus on the adaptation interface design aiming at obtaining the better adaptation efficiency.

The paper is structured as follows. In Section II the scenario of the study is discussed including the description of the scheme of the adaptation interface. Section III describes the adaptation process and the possible formats of the optical

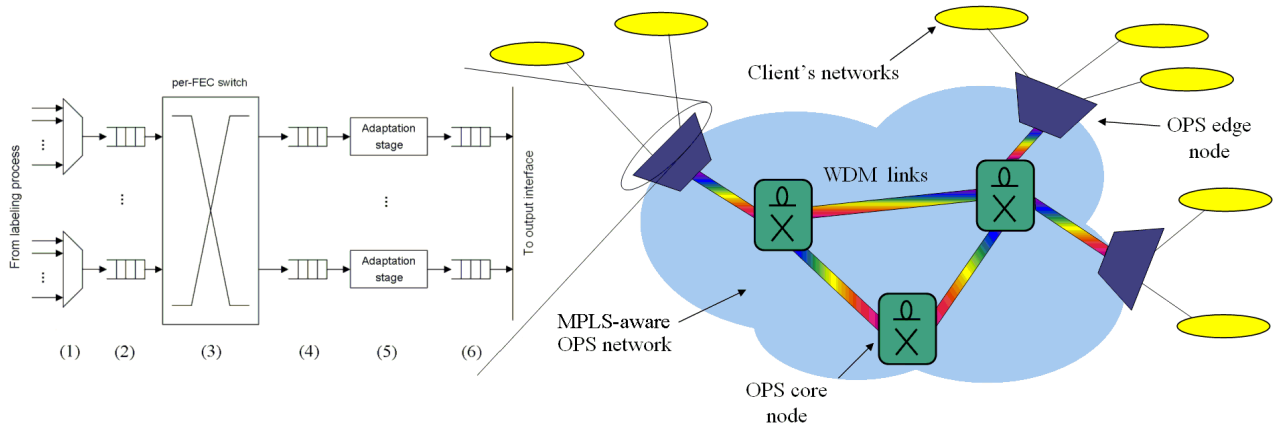

Fig. 2. Scheme of the adaptation interface

packet. Then in Section IV the simulation scenario and main numerical results are presented. Finally, Section V concludes the paper.

II. Scenario

We assume that the OPS network is MPLS aware [8], where the edge nodes of the optical network are responsible of the management of the MPLS label switched paths (LSPs) [9]. In our scenario, the edge nodes are also responsible of adapting the IP packets to the optical slots. As a consequence, an adaptation interface has to be included between the MPLS labelling process and the output interface (see Fig. 2).

A. Scheme of the Adaptation Interface

We assume that a labelling process labels the asynchronous, variable length packets coming from multiple sources such as LAN, IP router or switches and afterwards sends the labelled packets to the adaptation interface. Therefore, the role of the adaptation interface is to fit these packets into the corresponding synchronous, fixed-length slots, disregarding of output interface functionalities (such as E/O conversion or access protocols).

The adaptation interface consists of six blocks which perform the following functions (as shown in Fig. 2):

1) **Multiplexing**. The traffic coming from the labelling process is multiplexed into N flows.
2) **First FIFO stage**. The flows coming out from the multiplexers waiting the polling signal of the switching stage;
3) **Switching per FEC label**. The flows are forwarded according to their Forwarding Equivalent Class (FEC) label;
4) **Second FIFO stage**. The flows coming out from the previous switch are stored in FIFO queues waiting the polling of the following stage;
5) **Adaptation stage**. This stage performs the segmentation/aggregation functions in order to fit the client's datagrams into the correct optical packet format;
6) **Third FIFO stage**. Here the packets are stored with the optical format waiting the polling of the output interface;

It is important to notice that the internal speed of the adaptation interface has to be high enough to not create internal congestion. In such a way, the delay introduced by the adaptation only depends on the packetisation process (time spent waiting for an amount of information sufficient to fill one or more slots).

III. Adaptation Process and Packet Format Alternatives

For convenience, the client datagrams incoming from the electrical domain are named packets, while the term slot is reserved to the fixed-length packet of the optical domain.

A. Fixed-Length Packet Approach

The simplest adaptation process consists of inserting the packets directly into a sequence of slots. In this case, that we call *Fixed-Length Packets* (FLP) approach, the packets may be short enough to be carried by a single slot or has to be segmented into several slots, with some padding if needed. Once in the optical network, the slots will be treated as independent entities, each with its own header. At the outgoing edge of optical network, the packet will be reassembled.

Clearly, this is the quickest method since does not introduce additional delay on the packet processing. Nevertheless, from the performance point of view, the main problem of FLP approach is the inefficient resource utilisation since direct insertion may need considerable padding. For instance, if we consider slots of 1,000 bytes, a packet of 40 bytes (which represents the length of the majority of the IP packets [6]) only uses 4% of the slot resources. From this first observation, we can think that the shorter the slots, the lower the resource losses, i.e. the higher the efficiency. This is not true, since we need to consider that a packet longer than one slot has to be segmented and, as a consequence, a segmentation header has to be added in the slot. Moreover, it is necessary to introduce a synchronisation preamble and a guard-time

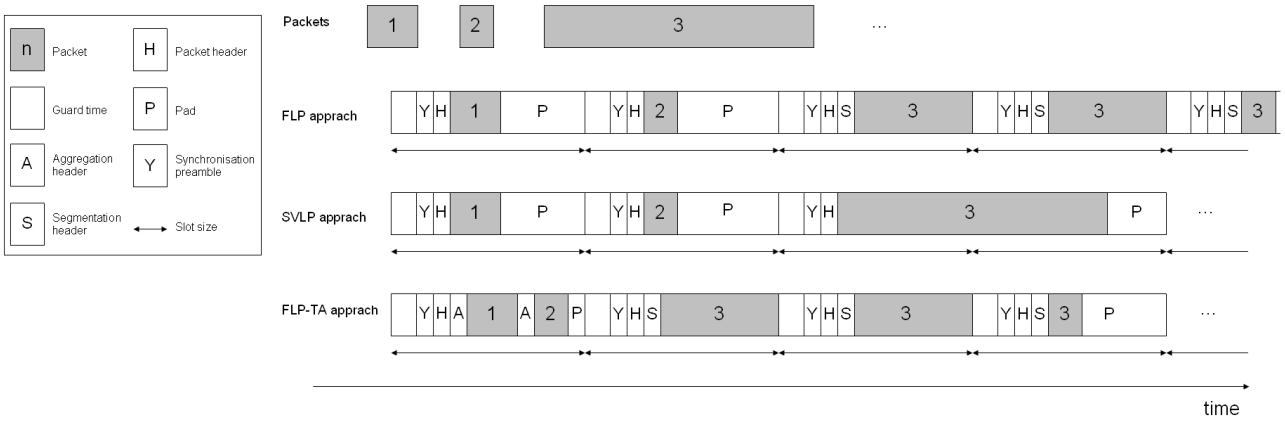

Fig. 3. Packet Format Alternatives

between two consecutive slots accounting for switching times of the constituent opto-electronic devices as well as packet position jitter [10]. This means that the shorter the slots, the higher the packet segmentation and therefore an higher number of no-strictly data information has to be included in a given period.

B. Fixed-Length Packet with Traffic Aggregation Approach

There are mainly two ways to increase the efficiency of the adaptation process. The first one is to introduce a *traffic aggregation mechanism* [12]. In this case, the adaptation stage is also in charge of aggregating the packets into the slots finding the maximum filling ratio at the expense of some additional delay. We call this method FLP-TA, and we consider the following simple aggregation mechanism.

The packets are organised and stored in different aggregation registers on per-FEC basis. When a new packet arrives the possible options are:

1) the packet size is shorter than the slot size, therefore the packet is not segmented. In this case the process controls the corresponding aggregation register:
   - if it is empty, the packet can be stored;
   - if it is occupied and its occupancy plus the size of the new packet is shorter than the slot size, the new packet is aggregated to the current contents of the register;
   - if it is occupied and its occupancy plus the size of the new packet is larger than the slot size, the current contents of the register is transmitted and the new packet is stored in the register.
2) the packet size is larger than the slot size, therefore the packet has to be segmented. In this case the process controls the corresponding aggregation register:
   - if it is empty, the new packet is segmented and each segment is transmitted. The register remains empty;
   - if it is occupied, the current contents of the register is transmitted and then the new packet is segmented and each segment is transmitted.

Finally, in order to limit the delay of the aggregation process, a time-out is used (one for each register). Each time a register is emptied and a new packet is inserted in the register, the corresponding time-out is initialised. If the time-out expires before the register is emptied again, the contents of the register is immediately sent, whatever is the current filling of the slot. This value is then the maximum delay that this approach can add to the packet processing at the adaptation interface.

C. Slotted Variable-Length Packet Approach

The second solution operates directly in the optical networking protocols and consists of the so called *Slotted Variable-Length Packets* (SVLP) approach [11]. This alternative recalls the train of slots model, where optical packets may have variable sizes as long as they are whole multiples of a slot. As in FLP, the packet may fit a single slot or span over multiple slots, except that different slots are now treated in the optical domain as a whole and then processed and sent sequentially. This approach allows the use of a single header for the whole train of slots and does not need a segmentation process allowing therefore overheads reduction. In contrast, the access protocol and switching functionalities may result more complex because they have to take into account the length of the train of slots.

Therefore, we have identified three different options, as illustrated in Fig. 3. Note that for this study, we assume the availability of optical switching matrices in the network nodes able to switch packets without any interferences between wavelength channels [4]. In this case, the guard time and the sync preamble are needed only between two consecutive train of slots, and not between two slots.

IV. Numerical Results

A. Simulation Scenario

We have set up a simulation scenario reproducing the above described adaptation interface in order to evaluate the efficiency of the FLP, SVLP, and FLP-TA approaches.

The incoming traffic of the adaptation interface consisted of 16 sources implemented using a self-similar traffic model

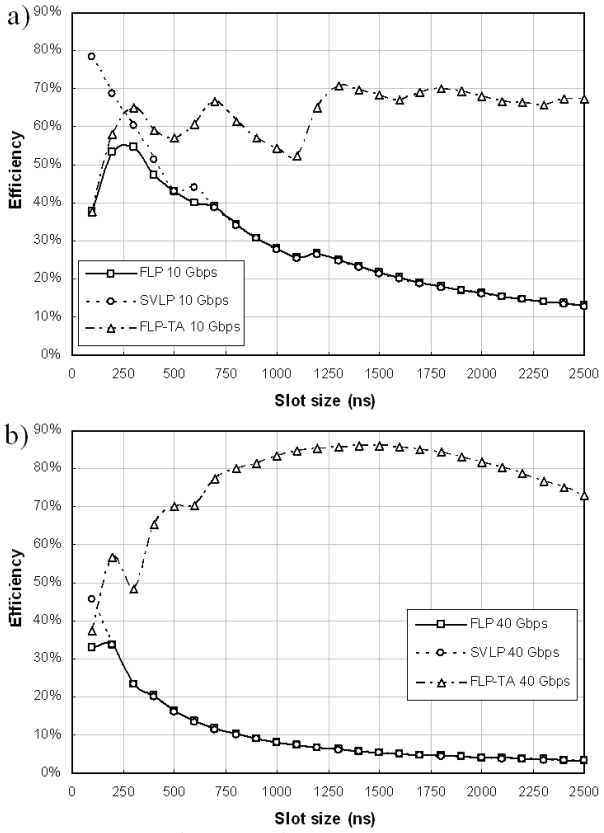

Fig. 4. Efficiency of FLP, SVLP, and FLP-TA approaches as a function of slot size (in ns) at a) 10 Gb/s, and b) 40 Gb/s

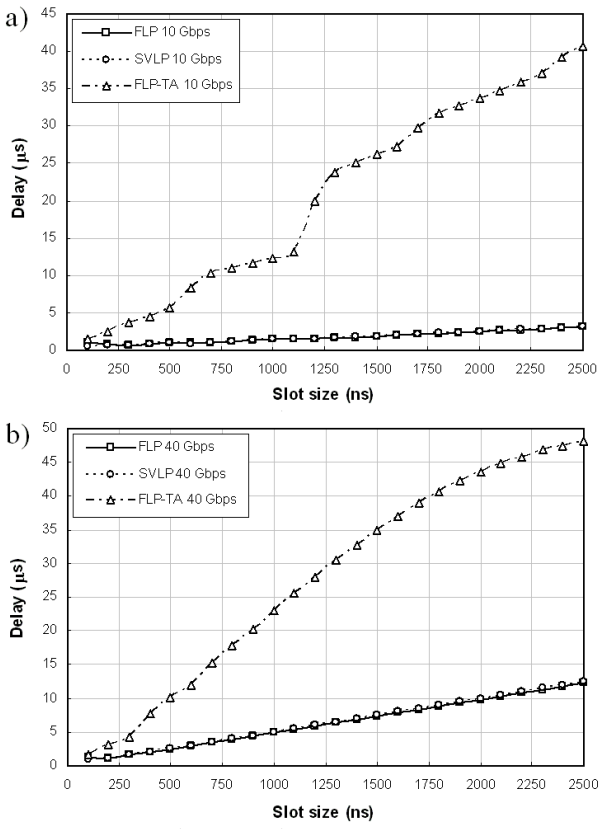

Fig. 5. Delay of FLP, SVLP, and FLP-TA approaches as a function of slot size (in ns) at a) 10 Gb/s, and b) 40 Gb/s

(the current more realistic traffic model) generating IP packets. Each source has been modeled as a superposition of 32 strictly alternating ON/OFF Pareto distribution sources with $\alpha = 1.2$, which leads to an Hurst parameter of $H = 0.9$ [13]. The IP packets have been generated according to the packet size distribution shown in [6]. We considered a guard-time of 50 ns (value imposed by the current available technology [3]), a segmentation header of 5 bytes, a packet header of 5 bytes, and a synchronisation preamble of 4 bytes. Three transmission bit-rates have been taken into account: 2.5 Gb/s, 10 Gb/s, and 40 Gb/s. For sake of simplicity, we assumed 30 FECs, and uniform traffic.

We want to remark that a different scenario may show different results (e.g. decreasing the number of aggregation queues may cause worst performance). The aim of this work is to illustrate some indicative values for further and more addressed studies.

B. Simulation Results

All the points of the following plots are steady-state values get from statistically significant measures obtained from the simulation results.

Figure 4 shows the efficiency of FLP, SVLP, and FLP-TA (time-out of 50 μs) approaches as a function of slot size considering a bit-rate of 10 Gb/s, and 40 Gb/s. We call efficiency the percentage of the slots used by transmitting the IP packets. At 10 Gb/s (Fig. 4a) the SVLP approach performs better results than FLP when the slot size is smaller than 700 ns. The curves overlap when this value is exceeded. In contrast, the FLP-TA approach performs better with slot sizes greater than 300 ns, reaching a 70% efficiency at 1300 ns. Note that the nonlinear behaviour of the curves is due to the IP packet size distribution. At 40 Gb/s, FLP and SVLP work in a similar way, while FLP-TA becomes strongly recommended with evident improvement of efficiency (86% at 1400 ns).

Figure 5 shown the average delay of FLP, SVLP, and FLP-TA (time-out of 50 μs) approaches as a function of the slot size considering a bit-rate of 10 Gb/s, and 40 Gb/s. This delay is the average time needed by the adaptation interface to transmit an IP packet to the output interface. For the FLP and SVLP approaches, this time is only due to the packet discretisation process. For FLP-TA, the aggregation process also contributes on the delay. From both figures (Fig. 5a and Fig. 5b) we can see that FLP and SVLP behave similar, while the aggregation process of FLP-TA adds a considerable delay; anyway it cannot exceed the time-out value. It is important to notice that the delay and the efficiency are strictly related: clearly, the higher the efficiency, the shorter the delay since the packets that waste less resource (i.e. better fitted into the slots) are also transmitted in less time.

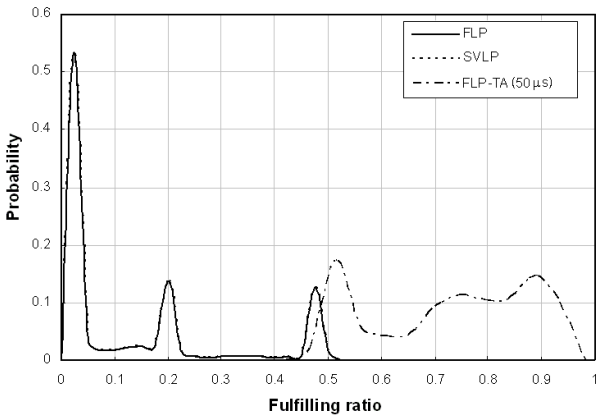

Fig. 6. Probability density function of the fulfilling ratio with 500 ns slot size at 40 Gb/s

TABLE I
EFFICIENCY AND DELAY OF FLP, SVLP, AND FLP-TA APPROACH AS A FUNCTION OF SLOT SIZE AND BIT-RATE

		2.5 Gb/s	10 Gb/s	40 Gb/s
FLP	Size (ns)	600	300	200
	Eff. (%)	69.5	54.7	33.6
	Delay (μs)	0.576	0.731	1.19
SVLP	Size (ns)	100	100	100
	Eff. (%)	89.5	78.3	45.7
	Delay (μs)	0.448	0.512	0.877
FLP-TA	Size (ns)	600	1300	1400
	Eff. (%)	74.0	70.7	86.6
	Delay (μs)	5.05	23.8	32.8

Figure 6 shows the probability density function of the fulfilling ratio of the slots at the output interface, comparing FLP, FLP-TA (time-out of 50 μs), and SVLP approaches at 40 Gb/s and with a slot size of 600 ns. This figure confirms the previous results showing that FLP-TA allows a high fulfilling ratio, with a majority of slots filled between 50% and 95%. In contrast, FLP and SVLP approaches are completely overlapped and cannot reach to fill more than 50% of the slots. This because the slot size results longer than the longest IP packet size distribution (1500 bytes at 40 Gb/s last 300 ns), hence the approaches behave in the same way. Therefore, since the approaches directly insert the packets into the slots without any modification, we can recognise the three peaks of the FLP and SVLP curves as the three main IP packet size probabilities (40, bytes, 576 bytes, and 1500 bytes [13]).

Table I compares the different approaches at different bit-rate considering the slot size where the highest efficiency is reached. From this table we can see that at 2.5 Gb/s the aggregation does not present evident benefits, and the SVLP approach performs better using a 100 ns slot size. The processing delay also confirm this result. Increasing the bit-rate to 10 Gb/s, the efficiency of all approaches decrease as well as the delay increase. Also in this case, the SVLP approach appears the better solution. At 40 Gb/s, FLP-TA outperforms the others, achieving 86.6% efficiency. It is worth noting that 100 ns is the *optimal* size for SVLP at any bit-rate.

The previous results show that the FLP-TA approach can achieve very good efficiency degree at expense of some additional processing delay. Two interesting results are now to know how the time-out value and the number of FECs parameters affect the efficiency of the FLP-TA approach.

For the former case, Fig. 7 shows the efficiency of the FLP-TA approach as a function of the time-out considering two slot sizes (600 ns and $1,200$ ns) at 2.5, 10, and 40 Gb/s. It can be seen that increasing the time-out value at 2.5 Gb/s, the aggregation does not increase the efficiency. On the other hand, the time-out value influences the efficiency at 10 and 40 Gb/s. Nevertheless, the curves show that as long as the time-out value is large enough, the efficiency flattens and does not change anymore (e.g. 20 μs for 600 ns slot size at 40 Gb/s).

Finally, Fig. 8 shows the efficiency of FLP-TA (time-out of 50 μs) as a function of the number of FECs. The slot in this case is 500 ns length. It can be seen that at 2.5 Gb/s the efficiency is independent of the number of FECs. In contrast, it has a strong impact at higher bit-rates, causing considerable efficiency degradation. At 10 Gb/s and 300 FECs, the efficiency goes down to less than 50%. At 40 Gb/s and 300 FECs, the value is about 45%. These bad results are mainly due to the fact that the time-out expires before the slots are completely filled. Therefore, a higher time-out value may increase the efficiency, but may also increase the delay.

V. CONCLUSION

The conclusion of this work can be summarised as follow. We have analysed the problem of adapting asynchronous, variable-length packets coming from the electrical domain to the synchronous, fixed-length packets used in the considered OPS network. We have suggested a scheme for an interface able to perform such adaptation process. In this process, the optical slot size becomes a crucial parameter, which strongly influences the bandwidth utilisation and determines the overall network performance. To find the optimum size, three different optical packet formats have been considered and their efficiencies have been evaluated.

Considering as real as possible traffic model, the bit-rate strongly influences the election of the *optimal* slot size. At 2.5 Gb/s, both FLP and SVLP give good efficiency for whatever slot size in the range of $300 \div 2,000$ ns. SVLP gives better results than FLP when the slot size is between 300 and 700 ns, while the approaches behave similar between 700 and 2000 ns. Increasing the bit-rate, the efficiency decreases notably both for the FLP and SVLP approaches. At these bit-rates, FLP-TA seems to overtake this problem. In this case, longer slots achieve better resources, but at expense of additional delay needed to fill the slots. Nevertheless, the performance of the FLP-TA approach strongly depends on the number of FECs. Indeed, increasing this value, the efficiency decreases considerably reaching worse values than SVLP approach.

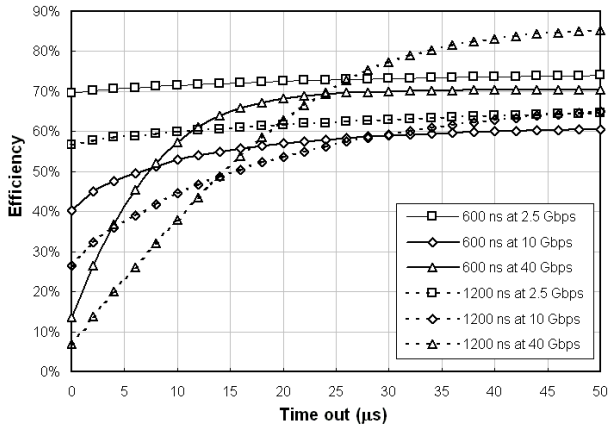

Fig. 7. Efficiency of FLP-TA as a function of the time-out considering 600 ns and 1,200 ns slot sizes at 2.5 Gb/s, 10 Gb/s, and 40 Gb/s

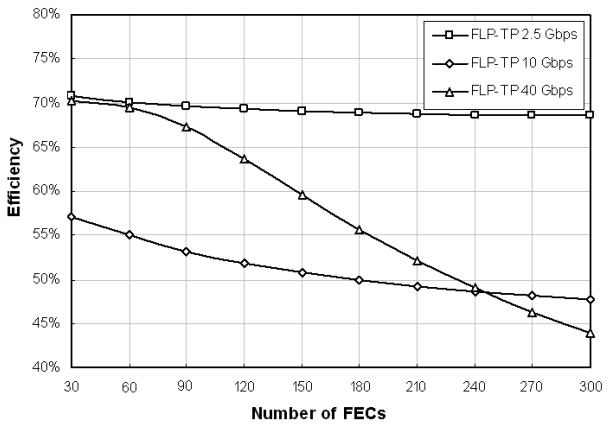

Fig. 8. Efficiency of FLP-TA (time-out of 50 μs) as a function of the number of possible destination nodes considering 500 ns slot size at 2.5 Gb/s, 10 Gb/s, and 40 Gb/s

Finally, we want to remark that the scope of this work was to give some indicative performance values of the FLP, SVLP, and FLP-TA approaches that can be used as guidelines for further investigations. For example, now it is possible to suggest another approach based on a combination of SVLP and traffic aggregation. In this case, the packets can be aggregated in the aggregation registers until the time-out expires or a maximum train of slots length is reached.

ACKNOWLEDGMENT

This work has been done within the DAVID project (IST-1999 11742) under contract with BTexact Technologies. This work has been also partially funded by COST 266 Action and MCYT (Spanish Ministry of Science and Technology) under contract FEDER-TIC2002-04334-C02-02.

REFERENCES

[1] B. Rajagopalan, D. Pendarakis, D. Saha, R.S. Ramamoothy, K. Bala, "IP over optical networks: architectural aspects", *IEEE Commun. Mag.*, vol. 38, no. 9, pp. 94-102, Sept. 2000.

[2] S. Yao, B. Mukherjee, S. Dixit, "Advances in photonic packet switching: an overview", *IEEE Commun. Mag.*, vol. 38, no. 2, pp. 84-94, Feb. 2000.

[3] L. Dittman *et al.*, "The IST project DAVID: a viable approach towards optical packet switching", to be published in *IEEE J. Select. Areas Commun.*

[4] P. Gambini *et al.*, "Transparent optical packet switching: network architecture and demonstrators in the KEOPS project", *IEEE J. Select. Areas Commun.*, vol. 16, no. 7, pp. 1245-1259, Sept. 1998.

[5] L. Tancevski, S. Yegnanarayanan, G. Castañ, L.S. Tamil, F. Masetti, T. McDermott, "Optical routing of asynchronous, variable length packets", *IEEE J. Select. Areas Commun.*, vol. 18, no. 10, pp. 2084-2093, Oct. 2000.

[6] K. Thompson, G.J. Miller, R. Wildes, "Wide-area Internet traffic patterns and characteristics", *IEEE Network*, vol. 11, no. 6, pp. 10-23, Nov. 1997.

[7] F. Callegati, "Which packet length for a transparent optical network?", in *Proc. SPIE Symposium on Broadband Networking Technol.*, Dallas, TX, Nov. 1997.

[8] F. Callegati, W. Cerroni, G. Corazza, C. Raffaelli, "Wavelength multiplexing of MPLS connections", in *ECOC 2001*, Amsterdam, Netherlands, Sept. 2001.

[9] E. Rosen, A. Viswanathan, R. Callon, "Multiprotocol label switching architecturee", *IETF RFC 3031*, Jan. 2001.

[10] C. Guillemot *et al.*, "Transparent optical packet switching: the European ACTS KEOPS project approach", *IEEE/OSA J. Lightwave Technol.*, vol. 16, no. 12, pp. 2117-2134, Dec. 1998.

[11] F. Callegati, W. Cerroni, "Time-wavelength exploitation in optical feedback buffer", in *Proc. Opticomm 2002*, Boston, MA, July 2002.

[12] G. Hébuterne, H. Castel, "Packet aggregation in all-optical networks", in *Proc. First Int. Conf. Optical Commun. and Networks*, Singapore, Nov. 2002.

[13] W. Willinger, M.S. Taqqu, R. Sherman, D.V. Wilson, "Self-similarity through high-variability: statistical analysis of Ethernet LAN traffic at the source level", in *Proc. ACM SIGCOMM 95*, Cambridge, MA, Aug. 1995.