Cluster Computing 7, 239–244, 2004

Performance Evaluation of Deflection Routing in Optical IP Packet-Switched Networks *

STEFANO BREGNI and ACHILLE PATTAVINA **
Politecnico di Milano, Department of Electronics and Information, Piazza L. Da Vinci 32, 20133 Milano, Italy

Abstract. In previous papers [5,6], an optical switch architecture was proposed to handle variable-length packets such as IP datagrams, based on an AWG device to route packets and equipped with a fiber delay-line stage as optical input buffer. Unfortunately, extensive simulations of optical networks built with switches of this type showed that considerable buffering capability would be required in order to achieve acceptable performance. In this work, therefore, we studied the effectiveness of packet deflection as a mean for solving packet contentions on outputs of optical switches. Optical transport networks were simulated, evaluating the performance of packet deflection routing, based on a traffic model adherent to real IP traffic measurements. Full-mesh and wheel network topologies have been considered, comparing results to assess deflection effectiveness. Our simulation results show that deflection routing leads to satisfying performance even using buffers with limited size. Furthermore, the average delivery delay does not suffer heavy penalty from packet deflection, even under heavy traffic conditions.

Keywords: Arrayed Waveguide Grating (AWG), Internet Protocol (IP), optical networking, optical switching

1. Introduction

The exponential growth of Internet users and the introduction of new broadband services have been fostering an unprecedented increase of network capacity. On the other hand, the IP architecture is being seen as the unifying paradigm for a variety of services and for the Broadband Integrated Services Network (B-ISDN), which has been foreshadowed since the 1980's. To face this challenge, considerable research is currently devoted to design IP full-optical backbone networks, based on Wavelength Division Multiplexing (WDM) technology, in order to relieve the capacity bottleneck of classical electronic-switched networks.

Photonic packet switching represents a potential solution [1–3]. Today, unfortunately, optical devices available on the market are still not mature enough to allow packet-by-packet operation in the optical domain. Optical burst switching has been proposed as intermediate solution between pure packet and circuit switching [4]. However, packet switching features an higher degree of statistical resource sharing, which should lead to a better bandwidth utilization when the network carries bursty traffic such as IP traffic.

This work is based on the optical switch architecture proposed in [5,6], based on an Array Wavelength Guide (AWG) to route packets to outlets and equipped with a fiber delay-line stage as optical input buffer. This optical switch was designed to handle variable-length packets, such as IP datagrams. Its performance was also evaluated for some typical statistical distribution empirically verified in the Internet.

The network architecture proposed in [5,6] simplifies the encapsulation of IP datagrams in optical packets by eliminating fragmentation issues. Moreover, it allows ultra-fine statistical resource allocation, being able to switch independently 40-bytes packets. Unfortunately, this switch would require considerable buffering to achieve acceptable performance, thus relying on expensive optical hardware and control electronics.

A possible solution, studied in this paper, is to implement efficient packet deflection inside the optical network, as a mean for solving packet contentions on outputs of optical switches. Thus, optical networks have been simulated to assess deflection effectiveness, based on a traffic model adherent to real IP traffic measurements. Full-mesh and wheel network topologies have been considered.

This paper is organized as follows. In section 2, the architecture of optical network studied in this work is introduced, summarizing the optical packet format and the switching architecture. In section 3, the system and traffic simulation models are described. In section 4, several simulations results are presented. Finally, section 6 draws some conclusions.

2. Architecture of the optical transport network

The general architecture of the optical network, as proposed in [5,6], is shown in figure 1 and consists of *M optical packet switching nodes*, each denoted by an unique optical address made of $m = \lceil \log_2 M \rceil$ bits, linked together according to a suitable topology. A number of *Edge Systems* (ES) interfaces the optical transport network with IP legacy electronic net-

* This paper is mainly based on the paper "Deflection Routing Effectiveness in Full-Optical IP Packet Switching Networks", by M. Baresi, S. Bregni, A. Pattavina and G. Vegetti, included in the *Proceedings of the IEEE Conference ICC 2003*, Anchorage, AK, USA (May 2003).

** Corresponding authors.
E-mail: {bregni,pattavina}@elet.polimi.it

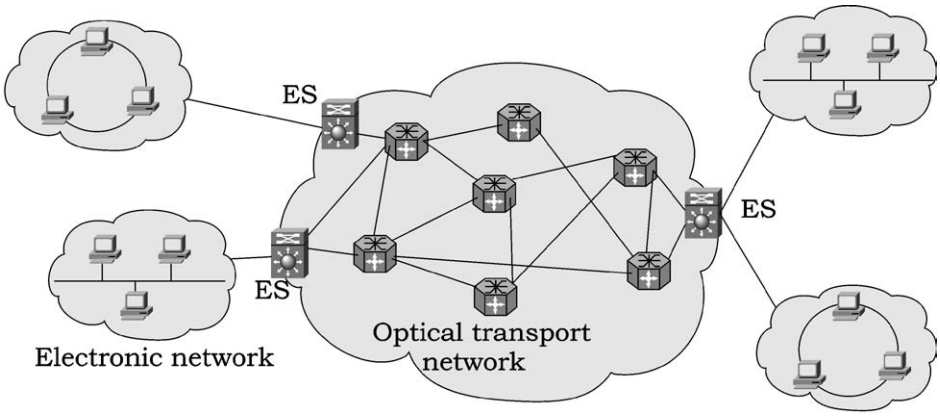

Figure 1. Architecture of the optical transport network.

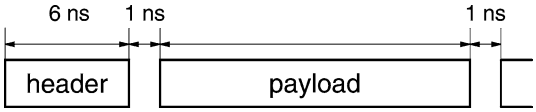

Figure 2. Optical packet format.

works. In our model, N ESs are connected to each optical node. Therefore, the total number of ESs is $N \cdot M$.

Edge systems multiplex IP datagrams from electronic networks and encapsulate them into optical packets with no fragmentation. Optical packets are then routed through the optical network to reach their destination ES, which delivers them to the destination electronic networks. The network operation is asynchronous: packets are transmitted between nodes without enforcing any time alignment. Conversely, internal operation within optical nodes is synchronous (slotted), to achieve lower contention probability [5,6].

2.1. Optical packet format

An optical packet is composed of a simple header, carrying the m-bits destination address, and a payload made of a single IP packet, as shown in figure 2. The optical header has fixed length while the payload size is not constrained.

The minimum time slot T of operation in optical nodes is the time needed for the smallest optical packet, carrying a 40-bytes IP datagram, to go from input to output ports. A 40 ns time slot seems appropriate, since 40 bytes are transmitted in 32 ns at the base speed 10 Gbit/s and 8 ns can be used for optical header transmission and to provide guard times. These are required between payload and header and between contiguous packets, to allow header processing and to account for some packet temporal skew inside switching nodes. Duration of guard intervals has been set to 1 ns. Therefore, header has duration 6 ns. It is transmitted at fixed rate 10 Gbit/s, while payload transmission can be set to higher rates, since the network is totally transparent to payload format and bit rates except for optoelectronic stages integrated into Edge Systems.

2.2. Optical switch architecture

The internal operation of optical nodes is synchronous, to achieve lower contention probability. Therefore, all packets entering input ports have to be aligned first to time slots, of duration $T = 40$ ns to accommodate smallest IP packets (40 bytes), before being routed by the switching fabric.

The structure of the optical switch is shown in figure 3. For a detailed description of its architecture and operation, the reader is referred to [5,6]. In this section, only its main features are highlighted.

Input WDM channels are demultiplexed, so that each wavelength enters the switch from a different inlet. At the switch output, W adjacent outlets, being W the number of wavelength per channel, are then multiplexed on the output WDM channel.

At the switch input, headers are first read and sent to the control electronics (H blocks). An n-stages *synchronization unit*, consisting of a series of 2×2 Semiconductor-Optical-Amplifier (SOA) switches interconnected by fiber delay lines of different lengths, aligns incoming packets to time slots.

The second stage is the *fiber delay lines (FDL) unit*, which stores packets to accomplish optical buffering and scheduling for coping with contention resolution on output ports. Tunable Wavelength Converters (TWCs) are used to route packets to the chosen delay line. The optical scheduling algorithm sets variable delays for packets entering the switching matrix. This algorithm even allows two packets entering the switching matrix in inverted order compared to that in which they entered the FDL unit, supposed that a sufficient maximum delay is available (buffer depth $D_{\max}$).

Finally, the third stage is the *switching matrix unit*, based on an Arrayed Waveguide Grating (AWG) device and two stages of TWCs, where the first stage is needed to route packets to the desired output and the second is responsible to convert the signal to a suitable wavelength, in order to avoid two packets to be transmitted using the same color.

2.3. Packet deflection

Packet deflection extends internal switch buffering, using network links as longer optical delay lines. Nevertheless, de-

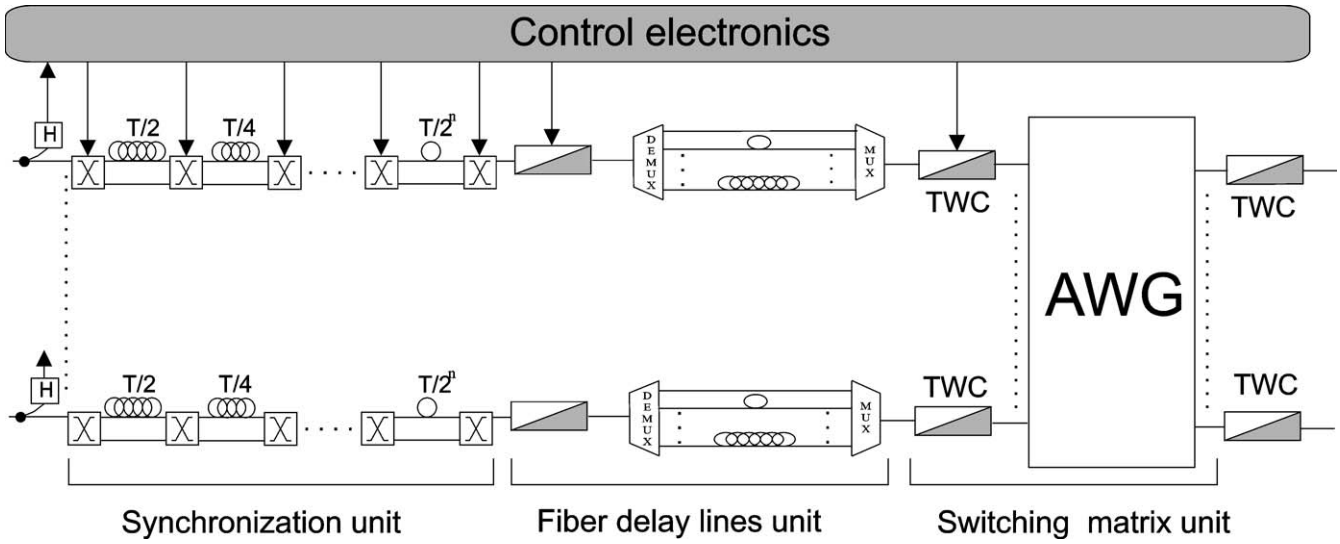

Figure 3. Structure of optical switch based on Arrayed Waveguide Grating (AWG).

flection generally leads to increasing network load. Thus, optimal deflection algorithms should direct packets to links scarcely loaded first, aiming at uniforming load among network links.

In this work, uniform packets deflection has been implemented: when a packet is deflected, it is routed with equal probability to one of the output links that are able to propagate it without further contention. At every switching node, deflected packets are handled as normal packets and are routed toward destination without any special processing. A *hop limit* H (i.e., *time to live*) is also enforced, to discard packets pinging too long inside the network.

3. Simulation model

According to the general network architecture shown in figure 1, we simulated the operation of optical transport networks for different network topologies and by varying the number of ESs, which generate and receive the IP traffic. Moreover, since the purpose of this work was to assess the performance of the transport network, we chose the simplest *star topology* to connect Edge Systems to optical switches.

In this work, we aimed at assessing the effectiveness of packet deflection in our optical transport network architecture. Therefore, we chose to simplify switch hardware complexity. In the FDL unit, we set the maximum buffering depth to $8T$. On the other hand, it has been shown in [5,6] that W should be set large enough in order to obtain satisfying performance, due to the *channel grouping* phenomenon. For this reason, the number of WDM channels used for the single input-output fiber has been set to $W = 20$.

For the aggregated traffic generated on each wavelength by the ES, we adopted a Poisson model, with interarrival times exponentially distributed. The length L of IP datagrams generated is a random variable, with empirical distribution according to real IP traffic measurements [7]:

$$\begin{cases} p_1 = P(L = 40\,\text{byte}) = 0.60 \\ p_2 = P(L = 552\,\text{byte}) = 0.25 \\ p_3 = P(L = 1500\,\text{byte}) = 0.15. \end{cases} \quad (1)$$

Hence, average packet length is 387 bytes. Moreover, the traffic pattern has been assumed addressed uniformly to all possible destinations of the network: therefore, the destination address of each packet is a random variable uniformly distributed between all possible ES addresses.

4. Simulation results

In this section, we present a selection of results obtained by the extensive simulations carried out. Full-mesh and wheel network topologies have been considered. Finally, results obtained in the two cases are compared.

All simulation results reported in this section are the central values of confidence interval estimates, with confidence level set to 95% and interval width on the order of 5%.

4.1. Full-mesh networks

We considered mesh networks with size $M = 3, 6, 9$. If not otherwise indicated, the number of Edge Systems connected to each transport switching node was set to $N = M - 1$. Hence, every switching node is connected to $M - 1$ ESs and $M - 1$ other switches. In this way, the traffic A [Erlang] offered by each ES equals the traffic offered on the average to each network link (network load). The packet hop limit has been arbitrarily set to a multiple of the network size M ($H = 0$, $H = 9$ or $H = 18$).

Figures 4 and 5 plot the packet loss probability, evaluated for networks with $M = 3$ and 6 nodes and hop limit $H = 0, 9$ and 18, versus the offered load A. The network exhibits better performance for higher levels of deflection. Conversely,

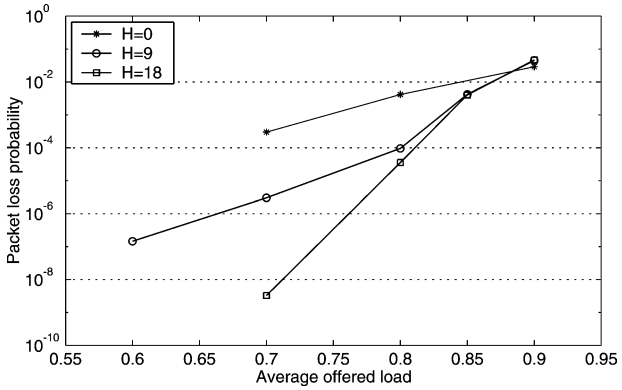

Figure 4. Packet loss probability in a full-mesh network with $M = 3$ nodes.

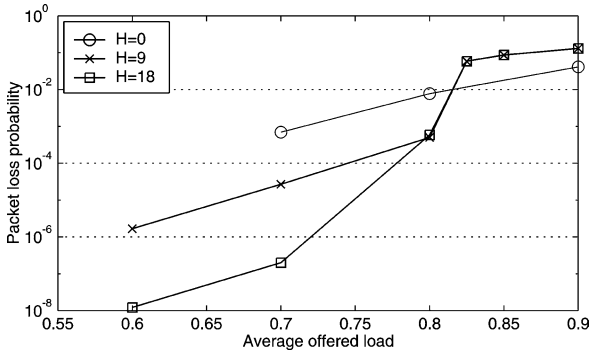

Figure 5. Packet loss probability in a full-mesh network with $M = 6$ nodes.

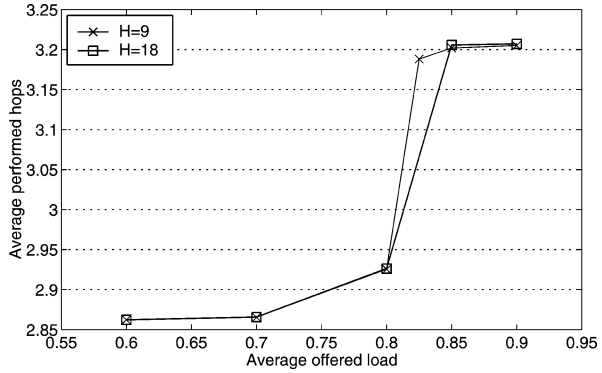

Figure 6. Average number of hops counted by delivered packets in a full-mesh network with $M = 6$ nodes.

under heavy traffic conditions, packet deflection worsens network performance. Deflected packets, in fact, represent a further load for single nodes, which leads to higher packet loss especially when heavy traffic is offered to the network.

The average number of hops required to deliver a packet is plotted versus the offered load A in figure 6, for $M = 6$ and $H = 9$, 18. Deflection routing does not increase drastically the average hop count. Even under congestion, a limited number of hops is sufficient to deliver packets in most cases.

In figures 7 and 8, networks with $M = 3$ and 6 nodes are directly compared, by setting the H parameter respectively to $H = 9$ and 18. We can see that loss probability increases as the number of nodes grows. This behavior is not determined by deflection routing, but is common to all switching systems

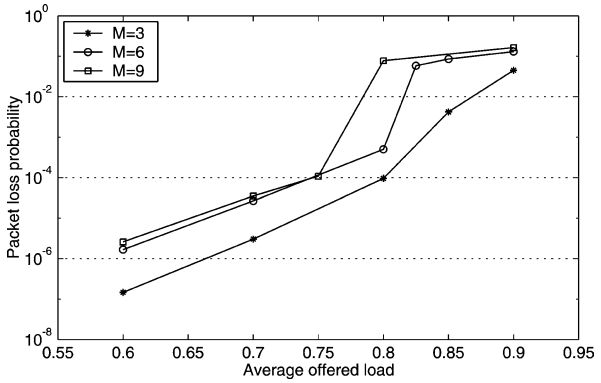

Figure 7. Packet loss probability in full-mesh networks with $M = 3, 6, 9$ nodes and $H = 9$.

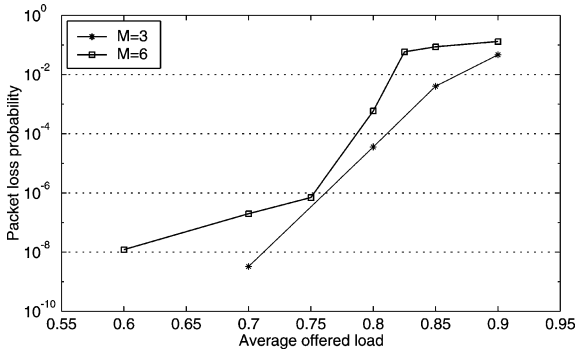

Figure 8. Packet loss probability in full-mesh networks with $M = 3, 6$ nodes and $H = 18$.

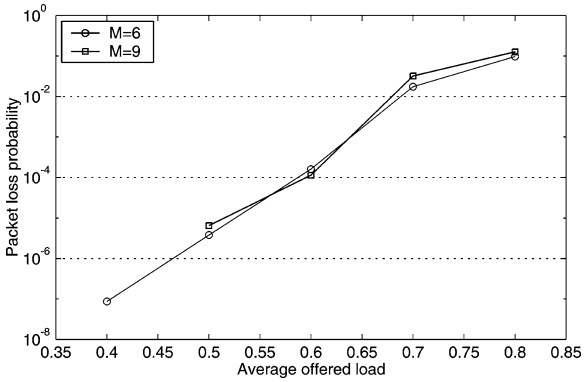

Figure 9. Packet loss probability in the network with nodes with $D_{\max} = 0$.

featuring input queuing (head-of-line blocking).

This consideration is supported also by figure 9, which shows the performance of networks with $M = 3$ and 6, where the maximum buffer depth has been set to $D_{\max} = 0$ (no input queuing) and the hop limit to $H = 18$. In these cases, the loss probability does not depend on the network size M.

To better understand network behavior under heavy load, we can examine the results shown in figure 10, where the loss probability versus the number of Edge Systems is plotted keeping constant the network load $A = 0.8$ Erlang.

Increasing the number of hosts per single transport node yields better performance. Since we are keeping network load constant, in fact, we are decreasing the traffic offered by sin-

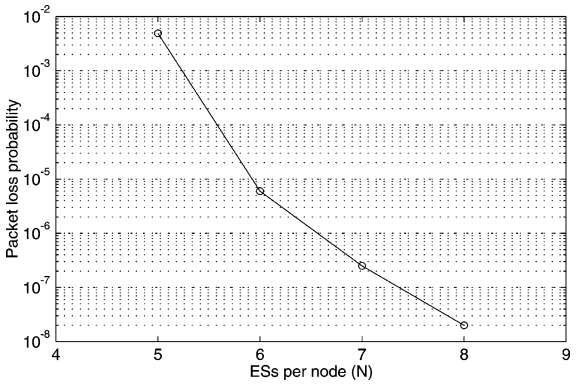

Figure 10. Packet loss probability in a full-mesh network with $M = 6$, $H = 18$, $N = 5, 6, 7, 8$ and constant offered load $A = 0.8$ Erlang.

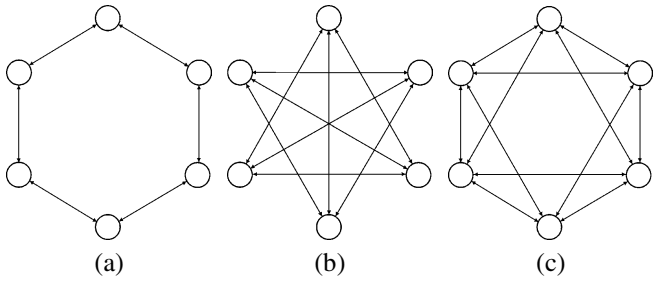

Figure 11. Wheel network topologies considered ($M = 6$).

gle ESs and therefore also the amount of traffic addressed to each ES. Thus, when a packet reaches the switching node directly linked to its destination, it has an higher probability to immediately delivered.

4.2. Wheel networks

In this section we present some results obtained for partially-meshed networks. Wheel networks are a particular class of regular network topologies that are easily represented placing nodes around a wheel. In particular, we considered 6-nodes wheel networks, with the three connection topologies depicted in figure 11. Table 1 summarizes the values of some characteristic parameters of these three network topologies.

The *connectivity factor* α is defined as

$$\alpha = \frac{2l}{M(M-1)}, \tag{2}$$

where l is the number of bidirectional links and M is the number of nodes. Therefore, α represents the ratio between the number of links in a wheel network and the number of links in a full-mesh network having same number of nodes.

The *network diameter* D is the maximum distance between two nodes. The *network order* Δ is the maximum number of links connected to a node. Finally, the *network number of hops* N_{H} is defined as the average distance seen from a node divided by the number of nodes of the network.

Figure 12 shows simulation results for these three kinds of network with a deflection limit $H = 18$ hops. Network performance worsens rapidly as α decreases. In fact, removing links from the network reduces deflection possibilities inside

Table 1
Characteristic parameters of the wheel network topologies shown in figure 11.

	l	α	D	Δ	N_{H}
Figure 11(a)	6	0.4	3	2	1.8
Figure 11(b)	9	0.6	2	3	1.4
Figure 11(c)	12	0.8	2	4	1.2

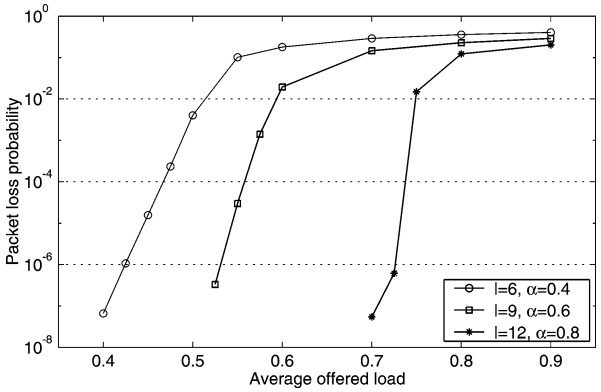

Figure 12. Packet loss probability in wheel networks with $M = 6$, $H = 18$.

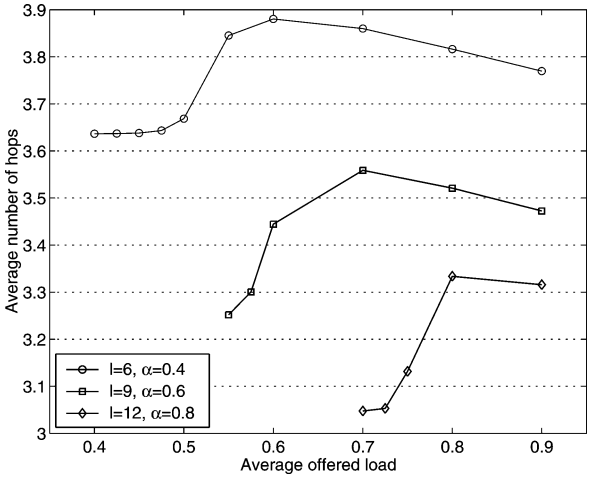

Figure 13. Average number of hops counted by delivered packets in a wheel network with $M = 6$, $H = 18$.

switches and this, combined with higher N_{H}, limits deflection effectiveness. The higher is the distance between two nodes, the higher is the chance for a packet to get deflected in the wrong direction around network topology, and thus also the probability to get lost.

Figure 13 displays the average number of hops performed by packets delivered to destination. All curves show a maximum: this is due to the fact that at very high loads many packets are discarded even before reaching the hop limit H, due to lack of available output links to any direction.

4.3. Comparison of full-mesh and wheel topologies

In figure 14, the performance of the full-mesh network is compared to that of the wheel network with $\alpha = 0.8$ ($M = 6$ and $H = 18$ in both cases).

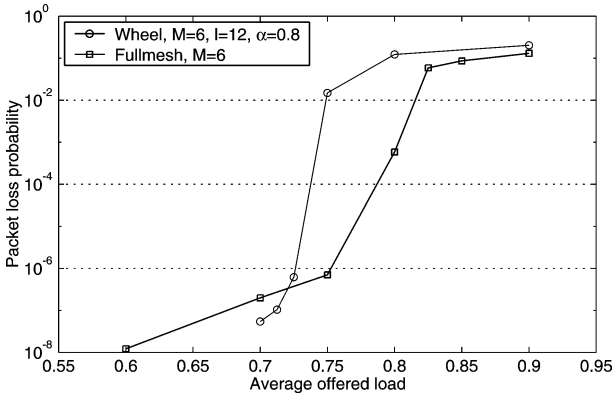

Figure 14. Comparison of full-mesh and wheel network ($\alpha = 0.6$) with $M = 6$ and $H = 18$.

Surprisingly, for medium-light loads the wheel topology outperforms the full mesh network. This behavior is explained observing that switching nodes have fewer input/output links in the wheel topology than in the full-mesh topology. Thus, input queuing may introduce a significant penalty.

5. Conclusions

In this work, we have studied the impact of packet deflection on the performance of optical IP packet switching networks with reduced buffer capacity. Based on the switch architecture proposed in [5,6], optical transport networks were simulated to assess deflection effectiveness, based on a traffic model adherent to real IP traffic measurements. Full-mesh and wheel network topologies were considered, comparing results to assess deflection effectiveness.

We have shown that, in full-mesh networks, deflection routing leads to satisfying performance even using buffers with limited size.

Furthermore, we pointed out that average delivery delay does not suffer heavy penalty from packet deflection, even in heavy traffic conditions.

Simulation results also confirmed that reducing the connectivity factor impairs substantially the performance of the optical transport network and the effectiveness of deflection routing. If the connectivity factor is low, cleverer deflection policies should be investigated, to avoid deflecting packets to nodes far from destination.

Acknowledgements

Work partially supported by the Italian Ministry of Education, University and Research (MIUR) under the FIRB project ADONIS.

References

[1] M.M. Renaud, F. Masetti, C. Guillemot and B. Bostica, Network and system concepts for optical packet switching, IEEE Communications Magazine 35(4) (1997) 96–102.

[2] D.K. Hunter and I. Andonovic, Approaches to optical Internet packet switching, IEEE Communications Magazine 38(9) (2000) 116–122.

[3] S. Yao, B. Mukherjee and S. Dixit, Advances in photonic packet switching: An overview, IEEE Communications Magazine 38(2) (2000) 84–94.

[4] C. Quiao, Labeled optical burst switching for IP-over-WDM integration, IEEE Communication Magazine 38(9) (2000) 104–114.

[5] S. Bregni, G. Guerra and A. Pattavina, Optical switching of IP traffic using input buffered architectures, Optical Networks Magazine 3(6) (November/December 2002) 20–29.

[6] S. Bregni, A. Pattavina and G. Vegetti, Architectures and performance of AWG-based optical switching nodes for IP networks, IEEE Journal on Selected Areas in Communications 21(7) (September 2003) 1113–1121.

[7] K. Thompson, G.J. Miller and R. Wilder, Wide-area Internet traffic patterns and characteristics, IEEE Network 11(6) (1997) 10–23.

Stefano Bregni was born in Milano, Italy, in 1965. He received his *Dott.Ing.* degree in telecommunications engineering from Politecnico di Milano. Since 1991, he has been involved in SDH transmission systems testing and in network synchronization issues, with special regard to clock stability measurement. Since 1999, he has been an Assistant Professor at Politecnico di Milano, where he teaches telecommunications networks.

He has been Senior Member of IEEE since 1999. He served on ETSI and ITU-T committees on digital network synchronization. He is author of the book *Synchronization of Digital Telecommunications Networks*, published by Wiley. He is Distinguished Lecturer on this subject of the IEEE Communications Society. He was Vice-Chair of the Transmission, Access and Optical Systems Committee of the IEEE Communications Society. He is Co-Chair of the *Access and Home Networks Symposium of the IEEE Conference ICC 2004* (Paris, France). He served in the Technical Program Committees of several ICC and GLOBECOM Conferences.
E-mail: bregni@elet.polimi.it

Achille Pattavina received the degree in Electronic Engineering (Dr.Eng. degree) from University "La Sapienza" of Rome (Italy) in 1977. He was with the same University until 1991 when he moved to "Politecnico di Milano", Milan (Italy), where he is now Full Professor. He has been author of more than 100 papers in the area of Communications Networks published in international journals and conference proceedings. He has been author of the book *Switching Theory, Architectures and Performance in Broadband ATM Networks* (Wiley). He has been Editor for Switching Architecture Performance of the IEEE Transactions on Communications since 1994 and Editor-in-Chief of the *European Transactions on Telecommunications* since 2001. He is a Senior Member of the IEEE Communications Society. His current research interests are in the area of optical networks and wireless networks.
E-mail: pattavina@elet.polimi.it