POLITECNICO DI TORINO
Repository ISTITUZIONALE

Multi-MetaRing Protocol: Fairness in Optical Packet Ring Networks

Original
Multi-MetaRing Protocol: Fairness in Optical Packet Ring Networks / Bianco, Andrea; Cuda, Davide; J. M., Finochietto; Neri, Fabio. - STAMPA. - (2007). (Intervento presentato al convegno IEEE International Communications Conference (ICC 2007) tenutosi a Glasgow (UK) nel June 2007) [10.1109/ICC.2007.394].

Availability:
This version is available at: 11583/1522483 since:

Publisher:
IEEE

Published
DOI:10.1109/ICC.2007.394

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

12 October 2024

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the ICC 2007 proceedings.

Multi-MetaRing Protocol: Fairness in Optical Packet Ring Networks

Andrea Bianco, Davide Cuda, Jorge Finochietto, Fabio Neri
Dipartimento di Elettronica, Politecnico di Torino, Italy
Email: *firstname.lastname*@polito.it

***Abstract*— We focus on Metropolitan Area Networks operating in packet mode and exploiting a single-hop wavelength division multiplexing (WDM) architecture. First, we briefly describe a specific slotted WDM optical network, based on a folded bus topology. Then, we address the fairness problem arising in this architecture and propose an extension of the MetaRing protocol to a WDM scenario. Two possible strategies are defined and analyzed. Finally, we show that both fair access and high aggregate network throughput can be achieved by properly handling node access through all WDM channels.**

I. Introduction

As Internet usage continues its growth, carriers continue to see a steady increase of packet data traffic in their Metropolitan Area Networks (MANs). Today's network solutions are mostly based on circuit-switched SONET/SDH rings that are not efficient in carrying data traffic, due to their inherent asymmetry, and bursty and self-similar behavior. Several evolutions of legacy SONET/SDH to packet-switched technologies are currently being proposed. For example, the IEEE 802.17 RPR (Resilient Packet Ring) standard aims at solving problems from which SONET/SDH networks suffer in supporting packet data by optimizing bandwidth sharing. However, as higher rates need to be supported, both SONET/SDH and RPR node costs increase, since all incoming/outgoing and in-transit traffic needs always to be processed electronically. Similar scaling problems arise in MAN infrastructures based upon switched Gigabit Ethernet, with additional concerns related to fair resource allocation and QoS control. Basically, in current solutions network scalability is limited because nodes must switch/process the full network bandwidth.

Due to advances in optical technology [1], new packet-switched networks can be devised that can sustain cost-effectively larger bandwidths. MANs seem to be one of the best arenas for an early penetration of these technologies. On the one hand, high capacity requirements can be satisfied by exploiting fiber bandwidth by means of Wavelength Division Multiplexing (WDM), without requiring node interfaces to access and electronically process the full network bandwidth. On the other hand, packet traffic can be handled by temporally sharing WDM channels, either by dynamically setting up *lightpaths* between nodes willing to communicate, or by exploiting statistical packet multiplexing in static channels.

In this context, single-hop optical ring networks operating in packet mode are considered a promising architecture for future MANs. The ring topology has been extensively proposed in the literature because of its simplicity and since it easily satisfies restoration requirements. Besides, the single-hop approach avoids complex switching in the optical domain and thus permits a cost-effective balance of optics and electronics. In these networks, nodes are equipped with few (typically one) transceivers, and each transceiver operates at the data rate of a single WDM channel. Paths between nodes are created by dynamically sharing on a packet-by-packet basis WDM channels, without requiring nodes to process the full network bandwidth. However, tunability at transceivers is required to exploit the fiber bandwidth by temporally allocating all-optical single-hop bandwidth between nodes in all available channels.

Due to the cost of tunability at transceivers, media access protocols that require packet-by-packet tunability only at one end of the all-optical path (i.e., either only at the transmitter, or only at the receiver) have been studied to save the cost of the still quite expensive tunable devices. Usually these protocols assume a *fastly* tunable transmitter and a fixed receiver, permanently tuned to a WDM channel [2]. When a node needs to send a packet, it simply tunes its transmitter to the receiver's destination wavelength. This implies that transmitter tuning times must be negligible with respect to the packet duration to obtain a good efficiency. Simple distributed access protocols can be designed for this tunable-transmitter/fixed-receiver architecture.

The scope of this work is to introduce in a WDM scenario fair access protocols based on the MetaRing concept [3]. We would like to remark that the design of these protocols in a WDM network imposes new challenges. Indeed, since nodes are typically equipped with one transmitter, the protocol should regulate not only traffic on each WDM channel but also access to different WDM channels to ensure good overall network performance.

II. Network Model

We consider a specific WDM optical packet network, physically made of two counter-rotating rings. This architecture was proposed and prototyped in the framework of the Italian national project named WONDER [4]. Each ring comprises N nodes and conveys W wavelengths, where typically $N > W$. Rings are used in a peculiar way: one ring is used for transmission while the other one is used for reception. To provide connectivity between the two rings, a folding point is needed, where transmission wavelengths are switched to the reception path, as sketched in Fig. 1. Transmitted packets travel from

1-4244-0353-7/07/$25.00 ©2007 IEEE

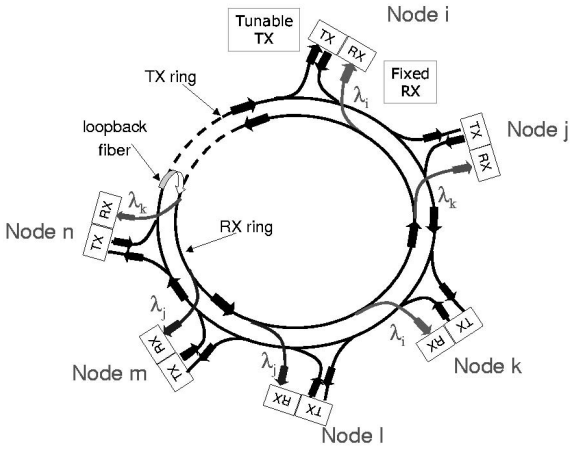


Fig. 1. Network architecture

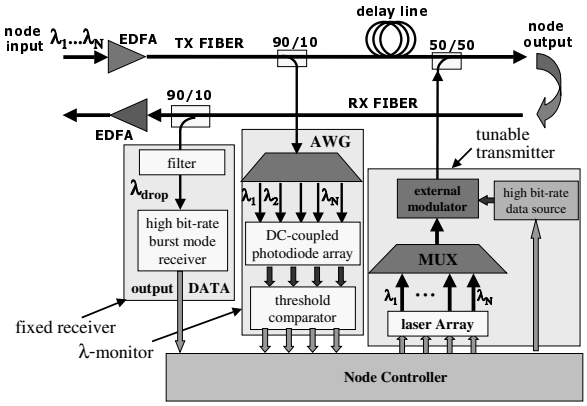


Fig. 2. Node architecture

the nodes towards the folding point in a first ring traversal, are switched to the reception path, and then received during a second ring traversal. If each node can become the folding point (i.e., if each node has a switching capability), then the network preserves the interesting restoration property of rings, as described in [5]. Although this architecture does not exploit wavelength spatial reuse, it avoids transmission impairment (e.g., noise recirculation) typical of ring topologies, while guaranteeing that all the network traffic accepted prior to the fault can be supported also after restoration, which may not be the case in ring networks with spatial reuse [6].

The network is assumed to be synchronous and time-slotted. The time a packet takes to traverse a whole ring is measured in time slots and it is referred as the Round Trip Time (RTT). During a time slot, at most one packet can be transmitted in one of the W available slots, one for each wavelength channel. Nodes are equipped with a single *fastly* tunable transmitter and a fixed receivers (see Fig. 2). Nodes exploit WDM to partition the traffic directed to disjoint subsets of destination nodes, each subset comprising the destinations whose receivers are tuned to the same wavelength. Nodes tune their transmitters to the receiver's destination wavelength, and establish a temporary single-hop connection lasting one time slot; due to the single transmitter architecture, a node can transmits only one packet per time-slot.

Access decisions are based on a channel inspection capability (similar to the carrier sense functionality in Ethernet – see [5]), by which nodes know which wavelengths were not used by upstream nodes in each time slot. From a design perspective, a suitable access protocol for the multi-channel network must avoid packet collisions while ensuring fairness together with high network throughput. A collision may arise when a node tries to transmit a packet on a time slot and wavelength which was previously used by an upstream node. This is avoided by giving priority to upstream nodes, i.e., to in-transit traffic, thanks to the λ-monitor capability (see Fig. 2).

A first level of fairness is achieved implementing an efficient a-posteriori packet selection strategy and exploiting the Virtual Output Queue (VOQ) structure [7]. While nodes in single-channel networks use a single FIFO (First In First Out) electrical queue, in a multi-channel scenario the FIFO queuing might lead to performance loss due to the Head of the Line (HoL) problem: a packet at the head of the queue might block other packets which could be transmitted on other channels. The HoL problem has been carefully studied, and it was demonstrated that it can be solved using the VOQ scheme described in [7] for Input-Queued switches, which permits to achieve 100% throughput under uniform unicast traffic. The basic VOQ idea consist of using separate queues, each one corresponding to a different destination, or to a different set of destinations (e.g. all the nodes receiving on the same wavelength), and properly selecting the queue which gains access to the channel for each time slot. In the case of the WONDER network, where usually there is more than one node receiving on the same wavelength, there is no difference between adopting a queue for each destination (N queues) or a queue for each channel (W queues); for simplicity and without loss of generality, we adopt the second solution.

III. MetaRing Protocol

A problem common to ring and bus topologies is the different access priority given to network nodes depending on their position along the ring/bus. Referring to Fig. 1, it is easy to see that an upstream node can "flood" a given wavelength, reducing (or even blocking) the transmission opportunities of downstream nodes competing for access to that channel, leading to significant fairness problems [2].

The MetaRing protocol was originally proposed to address fairness in ring networks where a single channel is available. In MetaRing, a control signal or message, called SAT (from SATisfied), is circulated from node to node in the opposite direction of data, possibly on a dedicated control channel. A node forwarding the SAT is granted a transmission quota Q: the node can transmit up to Q packets before the next

SAT reception. When a node receives the SAT, it immediately forwards the SAT to the upstream node on the ring only if it is satisfied, i.e., if

- no packets are waiting for transmission, or
- Q packets were already transmitted since the previous SAT reception.

If the node is not satisfied, the SAT is kept at the node until the node becomes satisfied. Thus, SATs are delayed by nodes suffering throughput limitations, and SAT rotation times increase with the network load. To be able to provide full bandwidth to a single node, the quota Q must be at least equal to the number of data slots contained in the ring: thus, $Q \geq RTT$. To avoid throughput limitation, nodes must be able to buffer at least Q packets: thus, the FIFO queue length must be larger than Q.

If a folded bus topology instead of a ring is assumed, as in the WONDER case, some issues need to be considered. First, the value of Q must be larger than in the ring case, since each time a SAT is forwarded, on average RTT slots are needed to reach the next node. Therefore, the quota Q must be at least N times the RTT to avoid starvation when only one node is active: thus, $Q \geq N \times RTT$. Queue lengths must increase accordingly.

Second, due to the folded bus topology, only the last node can potentially delay the SAT. Indeed, the last node has the lowest opportunities to access the channel since all other nodes are positioned upstream to it. As a consequence, when the first node forwards the SAT to the last one, all the other nodes have already renewed their quotas and typically have also began transmission. Thus, the last node receives the SAT and delays it until the channel becomes free. In overload conditions, the SAT is delayed until all nodes run out of quota. Since the SAT propagates in the upstream direction, each node releases the SAT and is able to transmit on average for RTT time slots until it is flooded by the traffic from the upstream node who has renewed its quota. As a result, when the SAT comes back to the last node, all nodes have a residual quota approximately equal to $Q - RTT$. Now, it is straightforward to realize that in the worst case, the SAT will be delayed at most for $N \times (Q - RTT)$ slots. Only when the last node exhausts its quota, the SAT is released and forwarded to the other nodes. However, since all these nodes are satisfied (i.e., they run out of quota), the SAT is simply forwarded with no delay until it reaches again the last node, where it is delayed again until satisfaction is achieved.

IV. Multi-MetaRing Protocol

To extend the MetaRing protocol to a multi-channel WDM network, we propose the Multi-MetaRing protocol that makes use of W SATs, where each SAT controls the traffic on a different wavelength channel. Metaring extensions to WDM rings were already proposed [8]. However, for the reasons explained above, folded bus topologies impose new challenges and previously proposed solutions cannot be directly re-used.

As in the case of a single channel network, the value of Q, for each channel, must be chosen so as to avoid node starvation even if a single node is transmitting on that channel: thus, $Q \geq N \times RTT$. Since there are W SATs circulating on the network, a node could delay more than one SAT to have equal opportunities to transmit on all channels. However, this would typically deteriorate network throughput: indeed, since nodes are equipped with only one transmitter, if more than one channel is blocked due to SAT retention, slots would be left empty. As a consequence, SAT retention policies must be defined; we consider two possible policies, named Hold-SAT and Release-SAT policies.

A. Hold-SAT Policy

The Hold-SAT (HSAT) policy states that nodes can retain (hold) more than one SAT. Thus, if a node receives a SAT while it is already delaying another SAT, it can retain this SAT if it is not satisfied on the corresponding channel. As a result, a node can hold up to W SATs; similarly to the single channel scenario, the last node will delay all SATs to access all channels. Under the HSAT policy, the Multi-MetaRing protocol is able to ensure absolute level of fairness in a short time window equal to the SAT rotation time, thus at most $N \times Q$ slots.

However, this policy can limit throughput performance, since the last node delaying more than one SAT when only one transmitter is available leaves some empty slots. To mitigate this problem it is important to establish the strategy that nodes must follow to schedule packet transmission. Typically, a *longest queue* strategy for selecting packets is adopted. If this is the case, a node retaining W SATs with queues approximately of the same length will delay all SATs until i) almost all queues are empty or ii) the quota on all channels is exhausted. As a result, the maximum delay experienced by all SATs is equal to WQ slots, while the maximum number of slots used for transmission is Q, resulting in temporal network utilization of $\frac{1}{W}$.

To overcome this problem, a *lowest quota* scheduling is proposed, with the aim of minimizing the time SATs are delayed. Under this strategy, the queue associated with the lowest residual quota is selected for transmission on empty slots. Thus, the queues associated to delayed SATs are served sequentially, allowing SATs to be forwarded as soon as possible. In overload conditions, one SAT will be delayed at most by Q slots, a second one by $2 \times Q$, and so on.

B. Release-SAT Policy

The rationale of the Release-SAT policy is that it is better to avoid to keep more than one SAT in a node to avoid the single transmitter blocking. As such, in the Release-SAT (RSAT) policy, a node can delay only one SAT at a time. Thus, if a node is already retaining a SAT when another one arrives, it forwards the lastly arrived SAT and renews its residual quota on the associated channel by increasing the available quota by a factor Q. In this way, priority is given to the already retained SAT, while forwarding with no delay other arriving SATs; thus, a node keeps a SAT until it is fully satisfied. In practice, SATs will be alternatively delayed by the last W nodes, who will

release SATs when satisfied. Although this solution improves network utilization by avoiding the retention of multiple SATs, it implies that nodes can cumulate a quota larger than Q, and that fairness can be achieved on longer time windows, that can be greater than $N \times Q$ slots. This fact implies that also node queues need to be larger, since they must temporally buffer packets on channels where quota is being cumulated.

V. Results and Analysis

In this section we present performance results obtained by simulation when considering a reference network with $W = 4$ wavelengths and a total of $N = 16$ nodes. The distance between two adjacent nodes is about 18km, i.e., 90μs; thus, the ring RTT is 1.45ms. Slots last 1μs, corresponding to a packet size of about 1250 bytes at 10 Gbit/s. Each node keeps W separate FIFO queues, one for each channel.

Two different traffic scenarios are considered: uniform traffic and unbalanced traffic. In the uniform traffic pattern, the whole capacity of the network is equally shared by all nodes. In the unbalanced traffic pattern, nodes are partitioned into two separated subsets: server and clients. The server subset contains only a single node, named *server*, positioned at the head of the bus to provide a worst case scenario. The server transmits at a high rate, equal to the capacity of one wavelength, with equal probability to the other $N-1$ nodes acting as clients. The remaining network capacity is shared by client nodes; each client transmits $\frac{1}{3}$ of its traffic toward the server and the remaining traffic to the other $N-2$ clients with equal probability.

A. Uniform Traffic Scenario

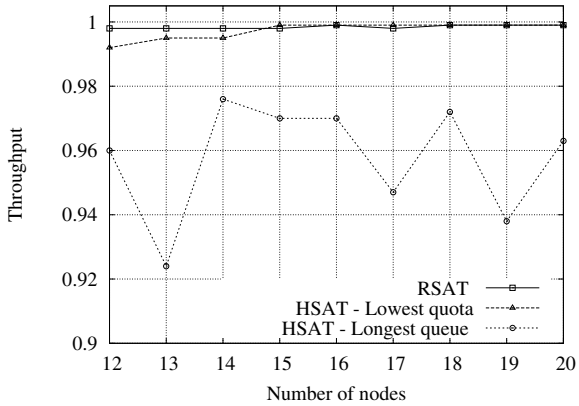
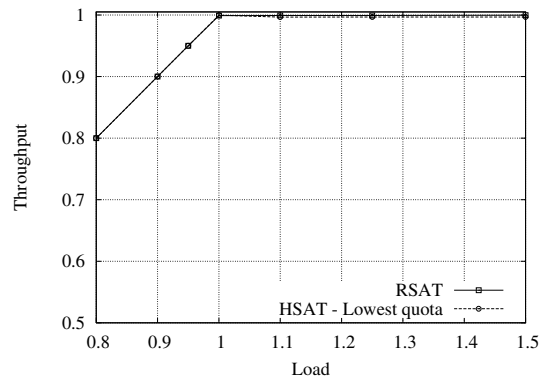
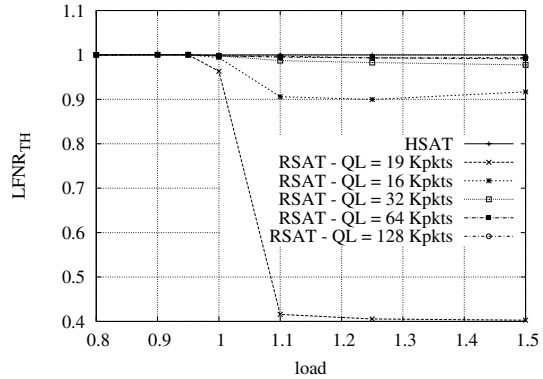


Fig. 3. Comparison of the throughput of the different strategies for different network sizes with offered load = 1.

We start considering the normalized throughput achieved by the Multi-MetaRing protocol under a uniform traffic scenario when the network load is equal to 1, varying the number of network nodes N. Fig. 3 shows how the RSAT strategy is able to achieve a throughput equal to 1, independently from the network size. On the contrary, the HSAT strategy performance depends on the scheduling policy. If a lowest quota scheduling strategy is adopted, performance is close to



(a)



(b)

Fig. 4. (a): Comparison of the throughput of RSAT and HSAT strategies; (b): Throughput fairness index for the RSAT and HSAT strategies.

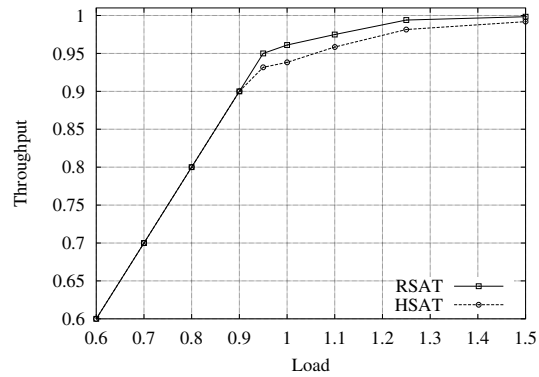
1 since the protocol is able to desynchronize the SATs when they are delayed by the last node. However, if the longest queue scheduling strategy is adopted, the achieved throughput deeply depends on the network configuration. Indeed, as the longest queue strategy equalizes queue lengths, all the SATs are released almost simultaneously by the last node. Hence, all nodes renew their quota on the different channels at the same time; as a consequence, nodes access the network in groups of W. Therefore, if the last group of nodes is composed by a number of nodes smaller than W, some channels are left empty and throughput drops. As an example, consider $N = 13$ and $W = 4$; in this case, the last node is left alone in the last group of nodes accessing the network. Thus, $3/4$ slots are left empty for the time SATs are delayed ($W \times (Q - RTT)$ slots). For this reason, in the remainder of the paper we omit to report the HSAT-longest queue results, and we refer to the HSAT-Lowest quota policy simply as to the HSAT policy.

Fig. 4(a) shows the throughput versus the offered load achieved by the RSAT and the HSAT strategies; both strategies achieve a throughput close to 1. Fig.4(b) shows the fairness index achieved by the protocols; we plot the ratio between

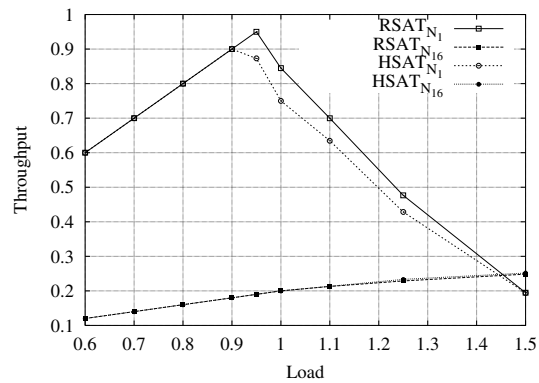
the throughput of the last node and the first one, labelled as the Last First Node Ratio (LFNR). For the RSAT strategy, we considered different queue lengths to illustrate the need of longer queues to achieve fairness. By definition the HSAT strategy is able to ensure absolute fairness in a single cycle, i.e., every $N \times Q$ slots: all nodes transmit Q packets on each channel, thus, a queue length equal to Q is sufficient to achieve fairness. On the contrary, if the RSAT strategy is adopted, the level of fairness deeply depends on the node queue length, which must be carefully selected as a function of the maximum achievable cumulated quota.

B. Unbalanced Traffic Scenario

We complete the analysis of the Multi-MetaRing protocol by considering its performance under an unbalanced traffic scenario. Fig.5(a) shows the normalized throughput for the RSAT and the HSAT strategies. The RSAT policy reaches a larger throughput when the network load is close to 1, while the HSAT strategy presents some loss due to the last node behavior. Indeed, the last node cannot stop more than one SAT; thus, the server can quickly renew its quota on the other channels, starting again to transmit on these channels, according to the MetaRing rules. On the contrary, when the HSAT policy is used, the last node can delay all SATs, blocking also the server. When the network is in deeply overloaded conditions, the HSAT performance approaches the RSAT one, since the network start behaving like under uniform traffic, according to the MAX-MIN fairness paradigm. Fig. 5(b) highlights the MAX-MIN fairness; as the network load increases, the server and the client throughput converge to the same value.



(a)



(b)

Fig. 5. MMR with RSAT policy: (a): Normalized throughput under unbalanced scenario; (b): Client/Server throughput under unbalanced scenario;

VI. Conclusions

We have discussed fairness issues arising in a WDM MAN network with N nodes and W wavelengths, based on a folded bus topology, where nodes are equipped with a single transmitter, a single receiver and W electronic queues. We have proposed two extensions of the MetaRing protocol to a WDM scenario, named RSAT and HSAT policies, which exploit W control signals, named SATs, to ensure throughput fairness on channel access.

The RSAT policy proved to be able to achieve the best performance both under uniform and unbalanced traffic scenarios; however, it presents throughput unfairness if node queues are not large enough. Indeed, nodes must be equipped with a large amount of memory (many times the value of the quota Q) to be able to ensure an acceptable level of fairness. On the contrary, the HSAT policy is able to ensure fairness in a relative short term cycle ($N \times Q$ time slots), with shorter queues, and achieves performance comparable with the one of RSAT if using a lowest quota scheduling among queues. For these reasons, the HSAT Multi-MetaRing seems the best candidate to control throughput fairness in the WDM network under study.

Acknowledgment

This paper was funded through the FP6 European Network of Excellence e-Photon/ONe.

References

[1] S. Yao, B. Mukherjee, and S. Dixit, "Advances in Photonic Packet Switching: an Overview," *IEEE Commun. Mag.*, pp. 84-94, Feb. 2000

[2] M. Ajmone Marsan, A. Bianco, E. Leonardi, M. Meo, F. Neri, "MAC Protocols and Fairness Control in WDM Multi-Rings with Tunable Transmitters and Fixed Receivers," *IEEE J. Lightwave Tech.*, pp. 1230-1244, Jun. 1996

[3] I. Cidon and Y. Ofek, "MetaRing – A Full–Duplex Ring with Fairness and Spatial Reuse," *IEEE Transactions on Communications*, vol. 41, no. 1, pp. 110–120, Jan. 1993

[4] The WONDER Project : http://www.tlc-networks.polito.it/wonder/

[5] A. Carena, V. De Feo, J. M. Finochietto, R. Gaudino, F. Neri, C. Piglione, P. Poggiolini, "RingO: An Experimental WDM Optical Packet Network for Metro Applications," *IEEE Journal on Selected Areas in Communications,* vol. 22, no. 8, pp. 1561-1571, Oct. 2004

[6] A.Bianciotto, J.M.Finochietto, R.Gaudino, F.Neri, C.Piglione, A.Samperi, "Fast and Efficient Fault-Recovery Strategies in the WONDER Metro Architectures", European Conference on Networks and Optical Communications, London (UK), July 2005

[7] N. McKeown, A. Mekkittikul, V. Anantharam, and J. Walrand, "Achieving 100% throughput in an input-queued switch", *IEEE Transactions on Communications*, Vol. 47, No. 8, pp. 1260-1267, Aug. 1999

[8] M.Ajmone Marsan, A.Bianco, E.Leonardi, F.Neri, S.Toniolo, "MetaRing Fairness Control Schemes in All-Optical WDM Rings", IEEE INFOCOM'97, Kobe, Japan, April 1997