SWING: Traffic Capacity of a Simple WDM Ring Network

Thomas Bonald, Sara Oueslati, James Roberts, Charlotte Roger

SWING: Traffic Capacity of a Simple WDM Ring Network

Thomas Bonald, Sara Oueslati, James Roberts, Charlotte Roger
Orange Labs, France Telecom, France

Abstract— **The paper presents a novel MAC protocol for a ring packet-switched WDM network. The protocol is based on both opportunistic access ensuring efficient utilization of slots and dynamic reservations that maintain a certain degree of fairness between stations. We show that, while a purely opportunistic access scheme is perfectly efficient, the impact of introducing the proposed reservation algorithm is limited in terms of lost capacity in any realistic traffic scenario.**

I. Introduction

There is renewed and growing interest in exploiting the high capacity and energy efficiency of dynamic optical switching to meet ever increasing Internet traffic demand. In this paper we consider a WDM ring network suitable for a metropolitan area. Our focus is on the design and performance evaluation of a novel MAC protocol that aims at keeping optical transport functions as simple as possible. It is called SWING for Simple Wdm rING.

The metro network consists of access and hub stations interconnected by two counter-rotating rings. Each station is an optical packet add/drop multiplexer (OPADM) receiving and delivering electronic data packets for connected users. These packets are transported within the network in optical packets each of which can contain several data packets. The network is multiservice and is intended to offer various QoS levels to meet the requirements of user applications. A fundamental design choice in SWING is to implement all QoS scheduling in the electonic domain preserving a simple transparent optical layer responsible only for transporting classless optical packets from one station to another.

A design objective for the optical transport layer is to maximize traffic capacity while ensuring a certain degree of fairness in the amount of bandwidth allocated to stations in congestion. To enable latency critical applications, it is important that each station experiences a reasonably smooth stream of available transmission opportunities requiring, therefore, a fine granularity fairness control. The optical layer must additionally be simple enough to easily accommodate anticipated technological developments in the OPADM field without significant redesign. Despite an abundant literature, we were not able to find an existing proposal that meets all these requirements.

The overall performance of the metro network obviously depends on the coupling between the electronic scheduling of data packets in the OPADM nodes and the formation of optical packets to be transported over the ring. Our present objective, however, is confined to investigating the traffic capacity of the optical layer alone. A key analytical result is to prove that a simple greedy opportunistic access policy, while unfair, is perfectly efficient in that the network is stable if and only if offered traffic is less than capacity on every ring segment. We then proceed to demonstrate by simulation that the distributed reservation mechanism of SWING does not bring excessive efficiency penalties.

The proposed WDM slotted ring architecture is described in the next section. We then present in Section II the optical transport MAC sub-layer and explain how slot reservation is used to impose fairness. The maximal efficiency of opportunistic transmission is proved in Section IV and the impact of the fairness mechanism is evaluated in Section V. Before presenting our conclusions, we briefly discuss in Section VI some related work on fairness in optical ring networks to explain why we are motivated to propose SWING as a simpler and more effective alternative.

II. Network architecture

We outline the architectural principles of the considered optical ring network and our approach to realizing QoS assurances.

A. A slotted WDM ring

The network is a time-slotted multi-channel ring where each slot on each wavelength can transport a fixed-size optical packet. There are N stations connected to the ring, numbered $1, 2, \ldots, N$ in the direction of the ring.

In addition to the data channels, the ring includes a separate control channel whose slots carry the information necessary for stations to know when they should add and drop optical packets on the data channels. Each control optical packet handles the corresponding slots on all data channels. The packet is demodulated, processed electronically and remodulated by each station in time for it to perform operations, as necessary, on the corresponding data slots.

The slot duration determines the optical packet size. It must be long enough to allow time for all control operations to be performed. On the other hand, it should be as short as possible to avoid assembly delays due to waiting for the arrival of a sufficient number of data packets. A reasonable time for control operations is 10μs for a 10Gbps channel leading to an optical packet size of 12.5KB, the equivalent of 8 Ethernet MTUs.

OPADMs may use either tunable or fixed transmitters and fixed receivers. In the evaluations reported below, we have

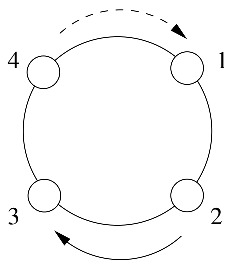


Fig. 1. Example of spatial reuse: station 2 may opportunistically transmit a packet to station 3 in a slot reserved by station 4

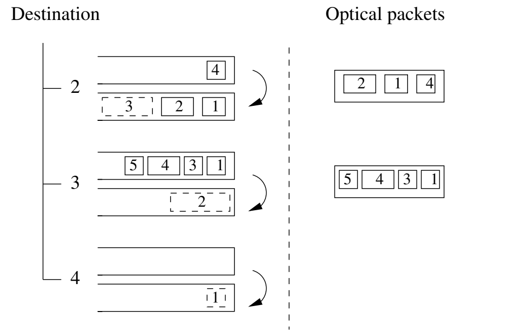


Fig. 2. Active queues at station 1 and optical packet creation

assumed all stations can access and receive all channels using fixed transmitters. This allows a station to send several optical packets in a single slot. The network is then equivalent in performance to a single channel ring with capacity equal to that of all channels combined.

Spatial reuse is an important feature of the present architecture. It increases the ring's performance in terms of traffic capacity, by allowing the same slot to be used for successive transfers along non-overlapping portions of the ring. This is enabled first by destination stripping: a destination (e.g., station 3 in Figure 1) immediately liberates the slot which can then be reused by the same station or by a downstream station (station 4 in the figure). In addition, reservations performed as described later are advertised by the control channel allowing, in the example of Figure 1, upstream station 2 to transmit a packet to station 3 using a slot reserved by station 4.

B. A two-tier MAC

Functionally, nodes implement two MAC sub-layers: *adaptation* and *transport*.

The adaptation sub-layer is responsible for creating optical packets. It electronically queues data packets (e.g., Ethernet frames) received from upper layer protocols in a set of per destination queues. The adaptation sub-layer initiates the process of seeking a transmission opportunity using a classical burst assembly algorithm (see [12], for example). Specifically, an optical packet is deemed to be ready for transmission to a given destination when it is full (i.e., there remains insufficient space for a further packet of size MTU) or when a timer expires indicating that some data packet to be included in the optical packet has been waiting longer than a certain threshold.

An optical packet is, in fact, only assembled at the last minute when the transport layer allows an emission. This allows a higher priority data packet to displace another packet whose arrival triggered the request to the transport layer. The transport sub-layer is described in detail in Section III.

To enable transparency in the optical layer, optical packets are not differentiated with respect to class of service. There are no priorities and no optical packet is displaced from its slot before arriving at its destination. Quality of service differentiation is realized by the adaptation layer that determines which data packets are included in successive optical packets, as discussed in the next sub-section.

C. Quality of service differentiation

We have not yet determined a definitive procedure for the formation of optical packets. It is rather a design option to allow maximum flexibility by designing a simple optical transport layer and confining QoS considerations to the adaptation layer where complex scheduling is easier to implement.

A regular stream of transmission opportunities (i.e., available slots) is provided to each station using the protocol described in Section III below. For each opportunity, the station decides on a destination and proceeds to assemble the optical packet from waiting data packets.

If the adaptation layer were designed to implement class of service differentiation, for example, the per destination queues would be structured in per class sub-queues. In Figure 2 we envisage the possible choice available to station 1 in the simple ring depicted in Figure 1. It has queues for premium and regular data packets for each of the three other stations. Numbers represent the arrival order of packets for each destination. In the imagined realization, optical packets to destinations 2 and 3 are filled first by premium packets and then completed by best effort packets if space remains (as for destination 2).

Alternative scheduling schemes could be implemented that are flow-based rather than class-based, for instance, or that realize fairness objectives for individual users connected to the metro network. This design space remains to be more completely explored. For present purposes, we assume SWING provides a solid basis on which to build an appropriate scheduling scheme at the adaptation sub-layer.

III. THE SWING PROTOCOL

SWING is intended to be as simple as possible while achieving high efficiency and a sufficient degree of fairness.

A. Opportunism and reservation

The protocol combines a distributed reservation scheme designed to ensure fairness while making maximal use of opportunistic transmission to avoid wasting capacity. Each data channel contains S data slots. The status of each slot, indicated by the control channel, is *free* or *busy* and *reserved* or *unreserved*. When *busy*, the control packet gives the destination, when *reserved* it also identifies which station has reserved the slot.

A specific slot, say slot 1, is used to synchronize the reservation scheme and defines the start of a cycle. The station reserves a slot for use in cycle c by marking the corresponding control packet as it passes in cycle $c-1$. The reservation is not performed for a specific destination since this is chosen at the last minute when the slot is actually used.

By opportunistic transmission we mean a station seizing every occasion to use a slot that is not busy to send a waiting optical packet to a destination that is not downstream of any station that may have previously reserved the slot. In normal traffic conditions, we expect the vast majority of packets to be transmitted opportunistically. Reservation is only necessary under congestion and, applying the distributed scheme described in Section III-B, ensures that all stations can then acquire a fair share of slots.

B. Slot reservation

The reservation scheme is fully distributed. Each station i maintains a counter R_{ij} of the number of slots reserved by station j since the last passage of the synchronization slot, for all $j = 1, \ldots, N$. Note that these counters are local visions of reservation status and, in general, are different from one station to another, due to possible preemption of the reservation by downstream stations, as further explained in the sequel.

Let X_{ij} be the number of packets waiting at station i for transmission to station j. We denote by $X_i = \sum_j X_{ij}$, the total number of packets waiting at station i. Station i seeks to reserve a slot on a given data channel when it has more packets waiting than the number of pending reservations, namely $X_i > R_{ii}$. This is realized by marking the control packet as reserved and entering its own identity as the reserving station. This reservation is not definitive. It can be pre-empted by a downstream station in the interests of fairness. The reservation of a slot held by station j is preempted by station i if i has outstanding waiting packets, that is $X_i > R_{ii}$, and fewer pending reservations than j, that is $R_{ii} < R_{ij}$.

This scheme guarantees any station suffering congestion can claim a fair share of ring capacity. It does not have the ambition to be perfectly fair in all conditions. Indeed, as previously noted the scheme is not usually effective in normal traffic conditions when opportunism is sufficient and more efficient. A station may also end up not reserving as many slots as it has waiting packets since it is not aware of preemptions. However, any residual packets at the end of a cycle are candidates for reservation during the next cycle, and so on, leading to progressively more aggressive reservation activity, as necessary to clear a backlog.

The following table presents the vision of each of $N = 4$ stations of the reservation status of the first 10 slots (initially unreserved) when all stations have many backlogged packets. The synchronization slot (slot 1) is initially at station 1 and it is assumed for simplicity that there are more than 10 slots between station 4 and station 1. Station 1 proceeds to reserve every slot. Station 2 preempts every second slot. Stations 3 and 4 preempt the slots as shown leading finally to a fair allocation to be exploited in the following cycle (as left by station 4).

Slot	1	2	3	4	5	6	7	8	9	10
Station 1	1	1	1	1	1	1	1	1	1	1
Station 2	2	1	2	1	2	1	2	1	2	1
Station 3	3	1	2	3	2	1	3	1	2	3
Station 4	4	1	2	3	4	1	3	4	2	3

TABLE I

RESERVATION STATUS OF THE FIRST 10 SLOTS GOING THROUGH STATIONS 1, 2, 3, 4.

It is clearly possible to generalize this preemptive mechanism to allow weighted fair shares, i.e., station i only preempts station j if $w_i R_{ii} < w_j R_{ij}$ for pre-assigned weights $w_1, \ldots, w_N$.

C. Packet transmission

A slot can be used by a station to transmit a packet in the following three cases.

1) The slot has been *reserved* by that station (that is, the reservation was made in the previous cycle and has not been pre-empted); this slot is necessarily free.
2) The slot is free and reserved by another station, but *spatial reuse* is possible.
3) The slot is *free* and *unreserved*.

Slots of any of these three types are seized whenever they can be used by some waiting optical packet.

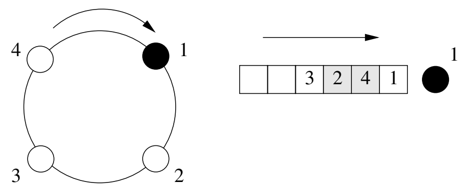


Fig. 3. Sequence of slots seen by station 1.

In Figure 3 we depict the sequence of slots seen by station 1. The numbers indicate the stations that reserved each slot. Shaded slots correspond to busy slots. Station 1 may transmit a packet in the first slot, because it holds the reservation (case 1). It cannot transmit in the following two slots, because they are busy. The fourth slot is reserved by station 3, but station 1 can transmit packets to stations 2 and 3 by spatial reuse (case 2). The fifth and sixth slots are free and unreserved and can hence be used by station 1 to transmit a packet to any destination (case 3).

D. Packet reception

When a station is the destination of a packet, the packet is extracted and the slot becomes free again. In particular, it can be immediately reused by this station, if permitted by the transmission rules described in §III-C.

Similarly, a station that holds a reservation must free the reservation at the passage of the slot. This slot can be immediately re-reserved by this station if permitted by the

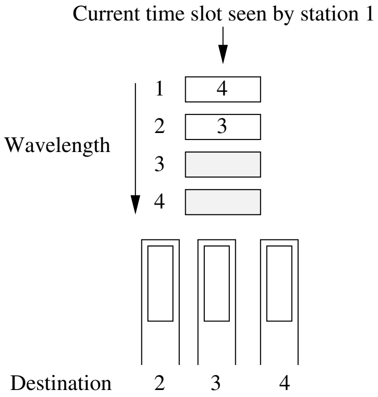


Fig. 4. Destination selection at station 1 for $L = 4$ data channels

reservation rules described in §III-B. Note that in many cases, the reservation will not be useful since most stations will have succeeded to transmit waiting optical packets opportunistically exploiting the last two cases of §III-C.

E. Packet scheduling

We consider three scheduling schemes to select the packet to transmit when several destinations are eligible: *round-robin* (RR) and *reverse round robin* (RRR), where stations are scanned cyclically in the orders $1, \ldots, N$ and $N, \ldots, 1$, respectively, and *longest-path first* (LPF) where the farthest station on the ring is selected.

To clarify the differences between these schedulers and to gain insight into their impact on resource allocation, we consider a simple example with $N = 16$ stations. Station 1 is assumed to have optical packets for all destinations. The current slot is free but has been reserved by station 8. First note that LPF selects destination 8 and this maximizes the slot utilization. Now let i denote the next destination to be served according to the RR and RRR schedulers. We must distinguish two cases:

1) if $i > 8$, RR selects destination 2, while RRR selects 8. Hence, RR tends to pick up the shortest possible path, while RRR optimizes spatial reuse.
2) if $i \leq 8$, both RR and RRR select i, which is suboptimal whenever $i < 8$.

In the case of multiple data channels, these are examined one by one following an arbitrary, predefined order. Figure 4 depicts the state of the current slots seen by station 1 for $L = 4$ data channels: slots on wavelengths 1 and 2 are free, but reserved by stations 4 and 3, respectively; slots on wavelengths 3 and 4 are busy. Again, all destinations are assumed to have waiting optical packets. Under RR scheduling, assuming station 2 is the next destination to be served, a packet for destination 2 is inserted on wavelength 1, then a packet for destination 3 is inserted on wavelength 2 (in both cases, spatial reuse is possible). The LPF scheduler selects a packet for destination 4 on wavelength 1, then a packet for destination 3 on wavelength 2.

IV. Efficiency of opportunistic transmission

In this and the next section, we focus on the traffic capacity of the ring network using SWING. This is defined as the stability limit beyond which the underlying queuing system, under stationary traffic, would cease to be ergodic.

For convenience, we index stations by integers modulo N. We refer to link l as that connecting station l to station $l + 1$. We denote by $[i, j]$ the set of stations between station i and station j, that is the set $\{i, \ldots, j\}$ if $i \leq j$ and the set $\{i, \ldots, N\} \cup \{1, \ldots, j\}$ otherwise. In this section we consider the capacity of a system where reservation is disabled.

A. Traffic characteristics

We assume that optical packets are generated according to a Poisson process for each source-destination pair. We denote by λ_{ij} the packet arrival rate from station i to j, with $\lambda_{ii} = 0$ for all i. The corresponding queue sizes X_{ij}, together with the state of each slot (occupation, reservation) and the location of the synchronization slot, form a Markov chain.

Note that the Poisson assumption is used only to make the system Markovian. We believe that the following results are insensitive to the statistics of the packet arrival processes beyond the arrival rates. Note, in particular, that the optical packet arrival process taking account of the burst assembly mechanism is typically less bursty than Poisson [12].

B. Stability condition

A necessary condition for stability is that the total traffic going through each link is less than the capacity of this link, that is:

$$\forall l = 1, \ldots, N, \qquad \sum_{i,j: l \in [i,j]} \lambda_{ij} \sigma < LC, \tag{1}$$

where σ denotes the common packet size, C is the channel bit rate and L the number of channels. These conditions give the maximum set of traffic intensities that the ring can support. We shall say that the MAC protocol is fully efficient if the ring is stable whenever these conditions are satisfied. It is not intuitively obvious but true that purely opportunistic access is fully efficient. This is stated in the following key result.

Theorem 1: In the absence of reservation, the ring is stable under the usual conditions (1).

Proof. We use the fluid limit approach of Dai [7]. Specifically, we denote by $\bar{X}_{ij}$ the fluid volume associate with the source-destination pair i, j. This is the number of packets X_{ij} after appropriate time-space scaling. The Markov chain is ergodic if the fluid model is stable in the sense that the total fluid volume $\sum_{ij} \bar{X}_{ij}$ empties in finite time, starting from any initial state such that $\sum_{ij} \bar{X}_{ij} \leq 1$.

Denote by $V_i = \sum_j \bar{X}_{ij} \sigma$ the fluid volume at station i. We define the fluid workload going through link l as:

$$W_l = \sum_{i,j: l \in [i,j]} \bar{X}_{ij} \sigma.$$

Let:

$$U = \max_{l=1,\ldots,N} W_l.$$

Since the fluid volumes $\bar{X}_{ij}$ are continuous functions of time, so is U. In addition, since $W_l \leq W_{l-1}$ for any link l such that $V_l = 0$, we have:

$$U = \max_{l:V_l>0} W_l.$$

When $V_l > 0$, station l uses all idle slots going through it so that link l is fully utilized. We deduce that:

$$\dot{W}_l = \sum_{i,j:l\in[i,j]} \lambda_{ij}\sigma - LC.$$

Let:

$$\delta = \min_{l=1,\ldots,N} \left(LC - \sum_{i,j:l\in[i,j]} \lambda_{ij}\sigma \right).$$

Note that $\delta > 0$ in view of (1). We have $\dot{W}_l \leq -\delta$ whenever $V_l > 0$. We deduce that at any time where $U > 0$ and U is differentiable, $\dot{U} \leq -\delta$. Finally, $U = 0$ after some finite time. The fluid model is stable. □

V. Efficiency of the reservation scheme

Reservations are necessary to enforce short-term fairness but decrease efficiency by limiting possibilities for opportunistic transmission. To evaluate this loss of efficiency we proceed by simulation.

A. *Efficiency metric*

Denote by ρ the system load, defined as the load of the most loaded link:

$$\rho = \max_{k=1,\ldots,N} \sum_{i,j:k\in[i,j]} \frac{\lambda_{ij}\sigma}{LC}. \qquad (2)$$

In view of Theorem 1, the stability condition for a purely opportunistic scheme is just $\rho < 1$. In the presence of reservations, the maximum load that guarantees stability is strictly less than 1. We refer to this quantity as the efficiency. This proves to be very sensitive to the considered traffic scenario.

B. *Traffic scenarios*

We consider $N = M+1$ stations, as illustrated by Figure 5 where $N = 5$. A particular station, say station 1, corresponds to the *hub* that connects the ring to the Internet. The other M stations are *access nodes* that connect users to the ring.

We assume access nodes are symmetric so that traffic is characterized by three parameters:

1) the traffic intensity from the hub to any access node,
2) the traffic intensity from any access node to the hub,
3) the traffic intensity between any two access nodes.

These parameters are given by λ_{12}, λ_{21} and λ_{23}, respectively. Since we are interested in the maximum sustainable load, it proves more convenient to characterize traffic through a single traffic intensity parameter, namely the total traffic $\lambda = M\lambda_{12}$ arriving at the hub from the Internet, together with the following two parameters characterizing incoming and outgoing traffic at any access node:

1) the ratio $\alpha = (M-1)\lambda_{23}/\lambda_{12}$ of *local* traffic (coming from any other access station) to *transit* traffic (coming from the hub);
2) the ratio $\beta = \lambda_{21}/\lambda_{12}$ of *upstream* traffic (going to the hub) to *downstream* traffic (coming from the hub).

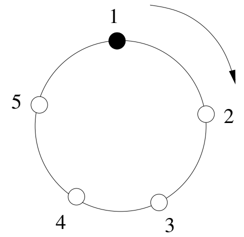


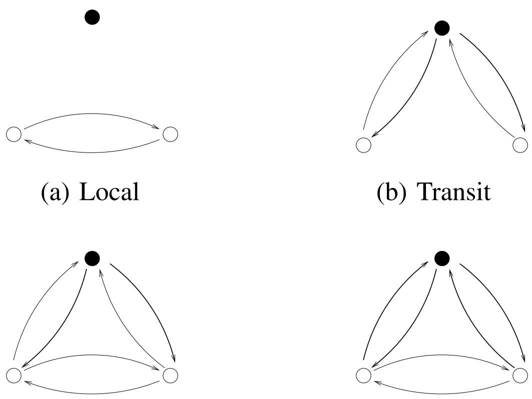
Fig. 5. Ring connected to the Internet through a hub (station 1, in black).

If $\beta \leq 1$, it may easily be verified that link 1 has maximum load. It then follows from (2) that:

$$\begin{aligned}\rho &= (\lambda + \lambda_{23}(1+2+\ldots+M-1)) \times \frac{1}{LC},\\ &= \left(1+\frac{\alpha}{2}\right) \times \frac{\lambda}{LC}.\end{aligned}$$

If $\beta > 1$, link N has maximum load and we get similarly:

$$\rho = \left(\beta + \frac{\alpha}{2}\right) \times \frac{\lambda}{LC}.$$



(a) Local (b) Transit

(c) Mixed (d) P2P

Fig. 6. Illustration of the traffic scenarios (hub station in black).

In all simulations, we shall consider the following four traffic scenarios, depicted by Figure 6:

(a) $\alpha = \infty$: all traffic is local;
(b) $\alpha = 0, \beta = 0.2$: all traffic transits via the hub and downstream traffic is preponderant;
(c) $\alpha = 0.2, \beta = 0.2$: traffic is a mix of local traffic and transit traffic, the latter being dominant;
(d) $\alpha = 0.2, \beta = 1$: traffic is also a mix of local traffic and transit traffic, but the intensity of transit traffic is equal in the upstream and in the downstream, which is typical of P2P traffic.

C. Measured efficiency

Unless otherwise specified, we consider a single data channel of $S = 100$ slots. The system is empty at the beginning of the simulation. Each simulation run lasts 10^4 cycles, that is 10s for a slot duration of 10μs. The system is considered as unstable if at least one station has more than 1000 packets in its buffer at the end of the simulation. We have verified that the simulation results are not very sensitive to this parameter.

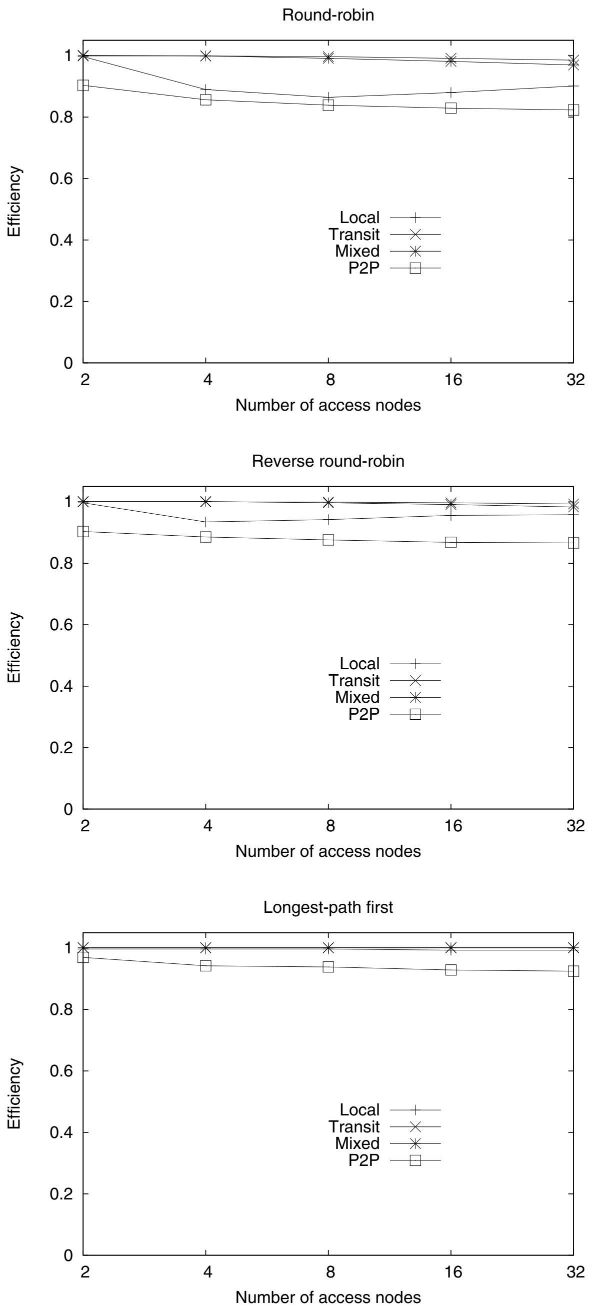


Fig. 7. Efficiency with respect to the number of access nodes M under round-robin, reverse round-robin and longest-path first scheduling.

Figure 7 gives the efficiency as a function of the number of access nodes, M, for the three considered scheduling policies: round-robin, where destinations are served in the order $1, \ldots, N$, reverse round-robin, where destinations are served in the reverse order $N, \ldots, 1$, and longest-path first, that always selects the farthest destination compatible with the reservation status of the slot in question.

We observe that efficiency is always higher than 80%. It is close to 100% for the transit and mixed scenarios, where link 1 is much more heavily loaded than other links. The worst scenarios are the local and P2P scenarios, for which all links are equally loaded. In these cases, the scheduling policy matters, with round-robin, reverse round-robin and longest-path first ranked from worst to best. For example, for $M = 16$ access nodes in the P2P scenario, efficiency is equal to 83%, 87% and 93% under round-robin, reverse round-robin and longest-path first, respectively. An example illustrating the reason for this ordering was given in §III-E.

Figure 8 shows the impact of the number of data channels L on efficiency for $M = 16$ access nodes under longest-path first scheduling. We observe that efficiency increases with L; it is close to 100% in all traffic scenarios for $L = 16$ channels. This can be explained by the fact that the number of slots is very large offering a lot of scope for opportunistic transmission. Reservation is hardly necessary and we fall back on the fully efficient scheme considered in Section IV.

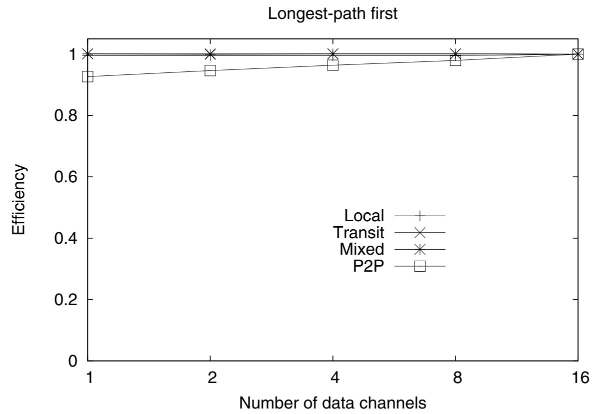


Fig. 8. Impact of the number of data channels for $M = 16$ access nodes.

D. Delay performance

Our main objective here is to explore the overall traffic capacity of SWING. Further evaluation is clearly necessary to characterize the fairness of the protocol and how this impacts performance perceived by the data flows whose packets fill the optical packets. However, the present set of simulation experiments can provide some indication of the likely delay performance and how this depends on scheduling.

Figure 9 gives the mean waiting time, averaged over all stations, for $M = 8$ access nodes and a single data channel. The time unit is taken equal to the cycle time $S \times \tau$, that is 1ms for $S = 100$ slots and a slot duration $\tau = 10\mu$s. We observe that the mean waiting time is very low until the load ρ approaches its maximum given by the network efficiency,

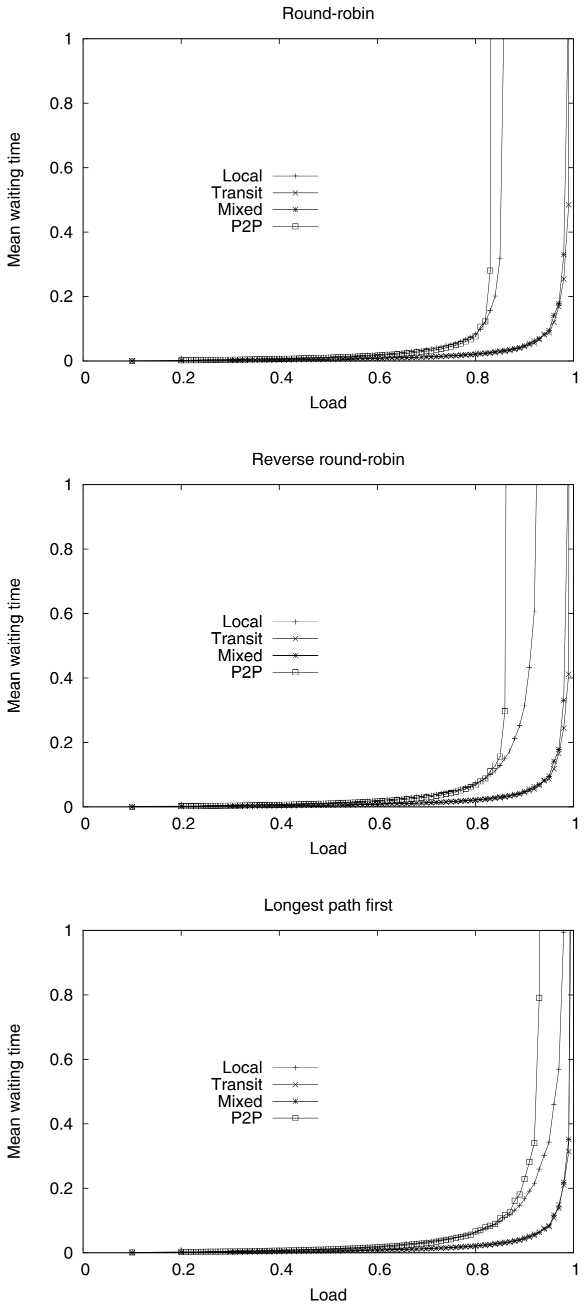


Fig. 9. Mean waiting time as a fraction of cycle time under round-robin, reverse round-robin and longest-path first scheduling for $M = 8$ access nodes.

cf. Figure 7. Under round-robin scheduling for instance, this maximum load is slightly higher than 0.8 for the local and P2P scenarios, where load is distributed over all links, and close to 1 for the transit and mixed scenarios, where load is concentrated on a single link. The reverse round-robin and longest-path first schedulers improve the delay performance in the worst traffic scenarios, namely local and P2P, due to higher efficiency.

VI. Related work on ring fairness

We are, of course, aware that there are many propositions in the literature for WDM ring network architectures with their associated fairness mechanisms. It is out of present scope to present a complete state of the art (see [9], for example) but it is important to explain why we have considered it necessary to propose SWING by highlighting some disadvantages of the principal alternatives.

A. Dynamic input rate control

The Resilient Packet Ring (RPR)[1] controls fairness by modulating the rate at which stations can insert packets [10]. Any station observing that it has insufficient ring capacity to satisfy its requirements is empowered to impose a limit on the input rate of a certain number of upstream stations. This limit is a locally calculated fair rate destined to avoid starvation without unduly depriving the other stations.

We have not sought to generalize this approach mainly because of its complexity and unproven efficacity. Considerable difficulty in defining the fair rate and specifying the conditions in which it should be imposed has been reported in [8].

B. Per station quota

Several proposals are based on the idea that each station is allowed to emit up to a certian quota of packets in a dynamically determined cycle. Fairness is ensured at the granularity determined by the size of the quota.

This principle was used in the Metaring protocol [6] and has been applied in the WDM ring architectures RingO [3] and Multi-MetaRing [2]. Fairness is realized by circulating a SAT (for 'satisfied') packet from station to station in the direction opposite to that followed by data packets. Stations release the SAT packet when they have emitted their quota or have no more packets to send. The quota is reset when the SAT packet makes a complete rotation. The problem with this approach is that the quota has to be large to ensure efficient utilization of the ring. This means there is typically a large interval between epochs when a station can emit a packet leading to potentially high delays.

The M-ATMR mechanism described in [1] is similar in principle to the SAT mechanism and has the same disadvantage since the quota must be large to minimize overhead. The M-FECCA mechanism described in [4] improves efficiency by allowing some opportunism but cannot reduce the inherently high fairness granularity.

SWING avoids using a quota mechanism to ensure any station can gain a fair share of ring capacity with maximum latency less than one cycle time.

C. Distributed queue control

The hybrid optical ring network, HORNET [11] realizes a different kind of fairness. It is argued that the MAC protocol is fair if it emulates an overall FIFO queue. HORNET implements a fairness algorithm called DQBR (for Distributed

[1] RPR is not WDM but uses a single wavelength; this does not invalidate the present discussion on fairness approaches.

Queue Bidirectional Ring) where a set of distributed counters is used to ensure upstream stations leave precisely the right number of free slots to be used by packets arriving earlier in downstream stations.

The FIFO emulation notion of fairness is not appropriate for our considered application of the WDM ring network. We ultimately seek to satisfy the QoS requirements of different applications and to realize a range of possible bandwidth sharing objectives. This requires controlled sharing of the ring bandwidth between its stations with a small fairness granularity to ensure low latency. A FIFO queue clearly does nothing to prevent some greedy user from monopolizing the resource leading to starvation and long delays for others.

VII. Conclusion

SWING is a simple optical transport protocol for a WDM slotted ring. It aims for maximum efficiency by exploiting all available opportunities for transmission while implementing a distributed reservation mechanism to ensure fair resource allocation in congestion. We have proven the maximal efficiency of greedy, opportunistic transmission and shown by simulation that the proposed fairness mechanism does not bring a significant penalty in lost traffic capacity.

While SWING is relatively simple, ensuring transparent and energy efficient optical transport, QoS for user applications relies on possibly complex scheduling in the electronic adaptation sub-layer where optical packets are formed. Our current research is focussed on analyzing the coupling between SWING and this multiservice, multi-user scheduling stage. We are confident that the fine granularity fairness of SWING is an adequate basis for the range of possible resource sharing policies implemented by the OPADM stations. We are also evaluating the impact of using a tunable emitter in place of the array of fixed emitters considered here.

Acknowledgment

This work has been performed with partial financial support from the French National Research Agency (ANR) through the project ECOFRAME.

References

[1] K. Bengi, H.R. van As, *QoS Support and Fainess Control in a Slotted Packet-Switched WDM Metro Ring Network*, GLOBECOM, vol. 3, November 2001, pp. 1494-1499.

[2] A. Bianco, D. Cuda, J. Finochietto, F. Neri, *Multi-MetaRing Protocol: Fairness in optical packet ring networks*, Proceedings of ICC 2007, 2007.

[3] A. Carena, V. De Feo, J. Finochietto, R. Gaudino, F. Neri, C. Piglione, P. Poggiolini, *RingO: An experimental WDM optical packet network for metro applications*, IEEE JSAC, Vol 22, No 8, Oct 2004.

[4] W.R. Chang, H.T. Wu, *FCCA: A Reverse-Direction-Based MAC Protocol Suite for WDM Metropolitan Ring Networks*, Proc. of Asia Pacific Conference on Communication, 2005.

[5] D. Chiaroni, *The French RNRT ECOFRAME project : packet technology perspectives in metro networks*, Invited presentation, APOC 2008, Octobre 2008, Hangzhou, China.

[6] I. Cidon, Y. Ofek, *Metaring: A full-Duplex Ring with Fairness and Spatial Reuse*,IEEE INFOCOM, June, 1990, pp. 969-978.

[7] J. Dai, On positive Harris recurrence of multiclass queueing networks: a unified approach via fluid limit models, The Annals of Applied Probability (5) 1, 1995.

[8] F. Davik, A. Kvalbein, S. Gjessing, *Improvement of resilient packet ring fairness*, IEEE GLOBECOM, 2005.

[9] M. Herzog, M. Maier, and M. Reisslein, *Metropolitan Area Packet-Switched WDM Networks: A Survey on Ring Systems*, IEEE Communications Surveys and Tutorials, vol. 6, no. 2, pp. 2-20, Second Quarter 2004.

[10] IEEE Standard 802.17, *Resilient Packet Ring*, September, 2004.

[11] I.M. White et al, *A summary of the HORNET project: a next generation metropolitan area network*,IEEE JSAC, vol. 21, number 9, November, 2003.

[12] X. Yu, J. Li, X. Cao, Y. Chen, C. Qiao, *Traffic Statistics and Performance Evaluation in Optical Burst Switched Networks*, Li, Journal of Lightwave Technology, Vol 22, pp 2722, Dec 2004.