Scheduling Issues in ECOFRAME Optical Packet Switched Ring

Bogdan Ušćumlić, Annie Gravey, Philippe Gravey, and Michel Morvan

***Abstract* — In the metropolitan area, traditional SONET/SDH circuit switched rings are likely to be replaced with optical packet/burst switching technologies. In this paper we consider a slotted WDM optical packet ring operating without resource reservation mechanisms. In such rings, optical packets in transit have priority over traffic to be inserted by the node. Packets to be inserted are thus queued according to their destination, in order to avoid head-of-line blocking. We focus on scheduling policies and compare several MaxWeight scheduling policies, including Oldest Packet First (OPF) which emulates FIFO queueing while avoiding head-of-line blocking. We show that there is a trade-off between implementation complexity and fairness, and identify the Largest Virtual Waiting Time First (LVWTF) scheduling policy as presenting both a low complexity and a good fairness performance.**

***Keywords* — fairness, MaxWeight, optical packet ring, scheduling, stability.**

I. Introduction

In this paper we consider an optical packet switching ring, designed for metro area, based on WDM technology and using both tunable lasers and Optical Add/Drop Multiplexers (OADM). This network is developed within the ECOFRAME project.[1]

In the past years, several projects have studied optical packet rings ([1]- [8]). They differ in terms of insertion and extraction methods, in terms of framing issues or in terms of architecture scenarios. In [9] it is shown that allowing any-to-any traffic in a single-wavelength, unidirectional optical packet ring can improve its capacity compared to a classical concentration/distribution scenario. The positive impact of WDM dimension on the performance of optical packet switched rings has also been studied in [10]. These studies, as the present one, assume that packets are inserted in an opportunistic manner, i.e. as soon as an appropriate slot is identified that can carry the packet to the destination. In other words, we assume that the network has been correctly dimensioned for the offered traffic and that no reservation mechanism is used.

In the present paper, we analyze the performance offered by several scheduling policies, in terms of stability, of implementation complexity and of fairness.

[1]This collaborative project has been funded by the French National Research Agency between 2007 and 2010.

Bogdan Ušćumlić is with the Mihajlo Pupin Institute, Volgina 15, 11050 Belgrade, Serbia; (e-mail: bogdan.uscumlic@pupin.rs).

Annie Gravey, Philippe Gravey and Michel Morvan are with the Institut TELECOM - TELECOM Bretagne, Brest, France; (e-mail: {annie.gravey, philippe.gravey, michel.morvan}@telecom-bretagne.eu).

The remainder of the paper is organized as follows. In Section II we give the basic characteristics of the network in study. In Section III we address the choice of suitable scheduling policy for an optical packet ring. The numerical results comparing different scheduling policies are given in Section IV. Finally, last section concludes the document.

II. The ECOFRAME Optical Packet Ring

The ECOFRAME Optical Packet Ring is a unidirectional network, where nodes are connected to the ring via Packet Optical Add-Drop Multiplexers, POADMs (Fig. 1), and to the client layers via an adaptation interface, responsible for multiplexing and demultiplexing client frames into optical packets.

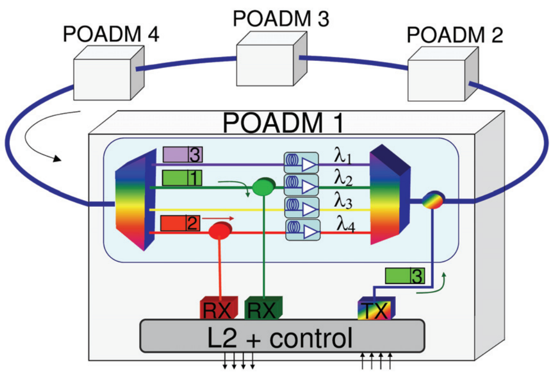

Fig. 1. ECOFRAME node architecture.

All the operations in the network are synchronized and occur at discrete time periods. Data is transported by using optical containers, so-called DATA packets, while control information is transported on a separate channel in control packets. Control packets report the occupancy status of the current time slot. A time slot on a wavelength can be either *busy* or *free*, depending on whether it carries a DATA packet or not.

Each node is equipped with a fully-tunable transmitter, able to dynamically change the transmission wavelength, from slot to slot and one or multiple wavelength receivers. An important feature of this architecture is that, although multiple wavelengths are used to carry data packets, a node is only allowed to insert at most one packet per time slot. In the general case, a node can receive packets on a fixed set of wavelengths. Received packets are queued in a reception node before being delivered to the client layers.

The ECOFRAME network is transparent to transit traffic. Packets to be inserted at a given node are queued till a free time slot is found on an appropriate wavelength.

In the present paper, we assume that for each pair of nodes, a single wavelength is used to carry all the traffic between the source and the destination nodes. However, the system is still WDM since many wavelengths can be used on the ring. It has been shown in [11] that this policy, although suboptimal in terms of network dimensioning, allows to consider simple per-destination FIFO queues accessing each a single wavelength instead of a complex set of queues accessing several wavelengths.

III. Scheduling in Optical Packet ring

Scheduling can be defined as the process of selecting a packet and a wavelength for insertion, by a ring node, per time slot basis. In the present case, scheduling consists in selecting, per time slot, a non-empty queue and a wavelength on which this queue should send its packets.

Contrarily to opaque systems where transit traffic is electronically handled in each node, and scheduling involves both new traffic and transit traffic, scheduling in ECOFRAME rings only affects new traffic since transit traffic is served transparently. A drawback of transparency is that in the general case, schedulers are “non-work-conserving”: if in a given time slot, queued packets cannot be sent on free wavelengths, no new traffic is served.

Transparency offered to transit traffic also impacts the stability of the ring. Indeed, only one packet can be served per time slot, and this packet can only be transmitted on a single wavelength. This implies that the scheduling policy has a significant impact on stability. This is a first point to be addressed.

A scheduling policy is deemed fair if all insertion FIFO queues within a given node present the same delay performance. This does not mean that end-to-end delays are identical since in particular transit delays depend on the respective positions of the nodes on the ring. Fairness is the second point to be addressed when selecting a scheduling policy.

Lastly, it is also important to assess the complexity associated with a given scheduling policy. Complexity is expressed here as the number of states and counters to maintain in each node. Stateless scheduling can be implemented when there is a single queue per node. Another extreme is observed when several counters per packet to be inserted have to be maintained.

A. Impact of scheduling on stability

We have chosen to mostly focus on *MaxWeight* scheduling disciplines [12]. MaxWeight scheduling guarantees the stability of different instances of generalized switch model, as first shown by Tassiulas & Ephremides in [13] and generalized by Stolyar in [12]. Dimensioning methods for ECOFRAME rings relying on MaxWeight scheduling policies are further studied in [14].

Simple and popular policies may not be MaxWeight, which leads to suboptimal dimensioning or to instability. We adapt here an example from [13] to show that Priority Queueing is non MaxWeight, and presents a smaller stable domain than MaxWeight policies.

Consider a ring with 2 wavelengths. Assume that a given node sends traffic from Queue1 on wavelength Λ_1 and from Queue2 on wavelength Λ_2; let $\lambda_1 = 0.5$ and λ_2 be the arrival rates to Queue1 and Queue2. We further assume that wavelength Λ_1 is always available, which we note with $p(\Lambda_1) = 1$, whereas a slot is available on wavelength Λ_2 with probability $p(\Lambda_2) = 0.5$.

According to [13], a necessary and sufficient stability condition for this system with two wavelengths, employing a “Longest Queue First” (LQF) scheduling policy, is given by the following set of inequalities: $\lambda_1 < p(\Lambda_1)$, $\lambda_2 < p(\Lambda_2)$, and

$\lambda_1 + \lambda_2 < p(\Lambda_1) + p(\Lambda_2) - p(\Lambda_1)p(\Lambda_2)$.

With the LQF scheduling policy, the above system is stable as long as $\lambda_2 < 0.5$.

Consider now a policy that implements priority to Queue1 (PQ). At each time slot, the node first attempts to transmit a packet from Queue1, and if there are no such packets, the node attempts to transmit a packet from Queue2. Obviously, under the PQ policy, the probability that the server attempts to serve Queue2 is only $(1-\lambda_1)p(\Lambda_2) = 0.25$, which implies that the system is only stable if λ_2 is smaller than 0.25, which is significantly smaller than the rate accepted by LQF.

This justifies our present focus on MaxWeight policies, for which stability conditions can straightforwardly be specified. According to MaxWeight policy definition from [12] even a popular scheduling policy, Round Robin, is not either of MaxWeight type, which is why it is not included in the analysis in this paper.

B. Considered Scheduling Policies

Each packet is characterized by its destination, and by the wavelength it should be transmitted to reach this destination. Therefore, we can classify packets according to:

1) either the destination (destination address queue),
2) or the destination wavelength (wavelength queue).

In the present case, we consider queues per destination. Each such queue is FIFO, i.e. we do not consider that insertion is class based within a queue. Note that such approach differs from the one taken in [15], for instance, where the authors consider separate queues for each (class of service, destination) pair. In our study, we try to reduce the number of queues, and thus to simplify the scheduling.

All scheduling mechanisms have to start by identifying, for each time slot, a set of eligible pairs (queue, wavelength). In a given slot, a pair is eligible if the wavelength is free, the queue is not empty and has to send packets on this particular wavelength. If the pair set is not empty, one eligible packet will be selected according to the rules that characterize the scheduling policy.

Four following scheduling policies are now considered:

1) *Random* (RAND): a queue to be served is determined by uniformly selecting one pair within the set of eligible pairs;

2) *Oldest First Packet* (OPF): the queue to be served is the one containing the packet that has experienced the maximum waiting time. This particular policy is FIFO within eligible packets and generally differs from a global FIFO which may suffer from head-of-the-line blocking.
3) *Longest Queue First* (LQF): the queue to be served is the longest queue in the set of eligible pairs;
4) *Longest Virtual Waiting Time First* (LVWTF): A modified LQF which weights the length of the queue by a factor inversely proportional to the flow rate used for dimensioning the system. LVWTF policy, for each node *i*, considers the variables $size_j/a_{ij}$ for eligible queues, where a_{ij} is the flow rate between nodes *i* and *j* (normalized to the wavelength capacity) and serves the queue for which this value is maximum.

The above scheduling policies, except RAND, are *MaxWeight* scheduling disciplines [12].

C. Scheduling Policy Performance Metrics

Scheduling policies are compared in terms of end-to-end queueing performance. End-to-end queueing latency is defined as the sum of insertion and extraction times. The insertion process leads to delays, because of the queueing process of data packets prior to insertion to the ring. The extraction time is due to the fact that a given node may receive packets on several wavelengths, whereas it delivers them to the client layers at a rate equal to a wavelength rate.

Fairness of a scheduling policy is defined by comparing the delay performance delivered by a given policy to the one delivered by the OPF policy. Indeed, OPF is the policy closest to a global FIFO, but not suffering from head-of-the-line blocking. The fairness metric considered below, although rather primitive, is both well suited to a comparison carried out by simulation, and nevertheless very significant.

Assume a set of flows $I_1, I_2, ..., I_n$. Let us note $D_{I_1}, D_{I_2}, ..., D_{I_n}$ and $L_{I_1}, L_{I_2}, ..., L_{I_n}$, the mean delays (in number of time slots) of packets belonging to flows $I_1, I_2, ..., I_n$, for OPF and some scheduling policy S, respectively.

The measure of the efficiency for the set of flows $I_1, I_2, ..., I_n$, when using the scheduling rule S is defined by:

$$\varepsilon_S = \sum_{i=1}^{n} | D_{I_i} - L_{I_i} | . \qquad (1)$$

The above metric is used in the next Section in order to compare the scheduling policies proposed here with OPF.

D. Scheduling Policy Complexity

As stated before, complexity is evaluated here by identifying the number of counters and states to maintain within a node in order to implement a particular scheduling policy. Hardware design is obviously much simpler when the number of counters to implement is fixed and if possible small.

Let us consider the complexity of the above scheduling disciplines:

1) RAND relies on a single occupancy indicator per queue.
2) In addition to the occupancy indicator, LQF and LVWTF necessitate maintaining a length indicator per queue.
3) In addition to the occupancy indicator, OPF necessitates storing arrival time for each packet.

Obviously, RAND has a very low complexity. LQF and LVWTF, although slightly more complex, necessitate a fixed number of counters (one per possible destination). On the other hand, the complexity for OPF is high since it necessitates a (potentially unknown) number of counters, each carrying the arrival time of a given packet. Although OPF would be a natural choice in terms of performance, since it closely emulates a global FIFO queue, we wish to assess whether a less complex scheduling policy would present a good fairness performance (i.e. is "close" to OPF in terms of delay performance).

IV. Assessment of Delay Performance

Using a ns-2 simulator developed for studying the ECOFRAME network, we have assessed the delay performance delivered by the four scheduling policies listed above.

All the simulation results in this work are given with a confidence interval of 10% (at confidence level of 95%). In order to assess fairness as defined in (1), the *zero measure of efficiency* is determined by $0_i = 0.1 \cdot D_{I_i}$. In other words, if $| D_{I_i} - L_{I_i} | \le 0_i$, then $| D_{I_i} - L_{I_i} |$ should be taken to be equal to 0, when calculating ε_S. The value of ε_S is then used to assess how close a given scheduling policy is to OPF.

A. Performance for a single-wavelength network

We first consider the simple case where a single wavelength is used by all traffic flows. We consider the network in Fig. 2 and we assess fairness in node A.

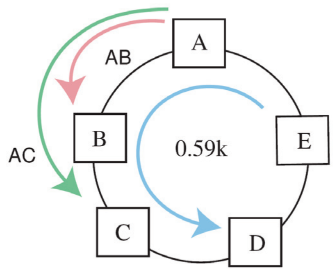

Fig. 2. Network used to assess fairness in node A.

Let us assume that nodes emit traffic flows as defined in scenario in Fig. 2, for $k = 1$.

Fig. 3 reports, for different scheduling policies, the average insertion times for flows AB and AC in two cases: when their rate is $A \rightarrow B = A \rightarrow C = 0.2$ and when their rate is $A \rightarrow B = 0.3, A \rightarrow C = 0.1$. Fig. 3(a) shows that when the rates for flows AB and AC are identical, the 4 mechanisms present identical behaviors.

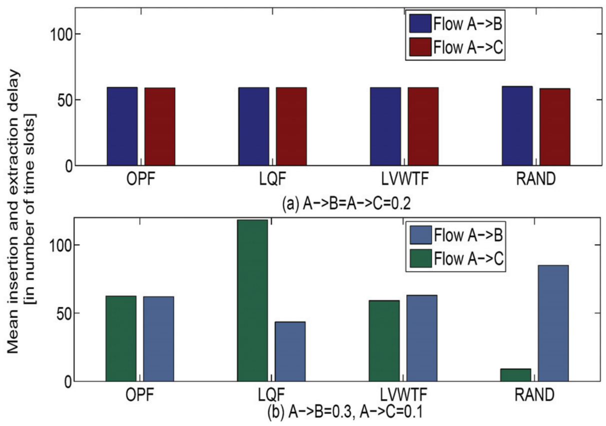

Fig. 3. Comparison of scheduling mechanisms in scenario from Fig. 2, for $k = 1$.

On the other hand, when the rates for flows AB and AC differ, the 4 mechanisms behave differently: some favor one flow above the other. Fig. 3(b) shows that OPF delivers the same insertion delay performance to the two flows, whereas LQF favors the larger flow (a result also observed in [16]). On the other hand, RAND favors the smaller flow, while LVWTF only slightly disadvantages the larger flow.

Table 1 presents the efficiency scores for different scheduling rules and different values of *k*. Table 1 shows that the deviation of the optimal performance increases with the value of *k*, for both LQF and RAND scheduling. This is not the case for LVWTF, which presents only a small deviation from OPF, for all values of *k*.

TABLE 1: EFFICIENCY MEASURE VALUE FOR LQF, LVWTF AND RAND.

	ε_{LQF}	ε_{LVWTF}	ε_{RAND}
k=0.2	0	0	0
k=0.4	0	0.27	0
k=0.6	0	0.65	0
k=0.8	1.40	1.46	0.88
k=0.9	5.77	2.27	4.74
k=0.95	14.66	2.51	12.03
k=1.0	74.41	0	76.33

As specified above, we have assumed here that each node sends traffic as specified in the dimensioning process.

This is not necessarily a plausible situation. In general, the dimensioning process considers “busy hour” estimates to dimension the network, and actual flow rates are lower than these estimates. In order to assess whether LVWTF is still close to OPF in this case, we consider an *under-loaded* scenario.

We consider the same traffic pattern as in Fig. 3(b), but with a difference that the flow AB is now being under-loaded, ie. $A \rightarrow B \leq 0.3$ holds. Fig. 4 shows the delay performance of packets inserted at node A, for OPF and LVWTF policies in function of AB rate.

Although the weighting coefficients that LVWTF policy uses to choose among the queues are not correct for flow AB, Fig. 4 does not reveal an important degradation in this policy performance. Moreover, LVWTF follows quite well the optimal performance reached by OPF for both destinations, with a result discrepancy of only several time slots.

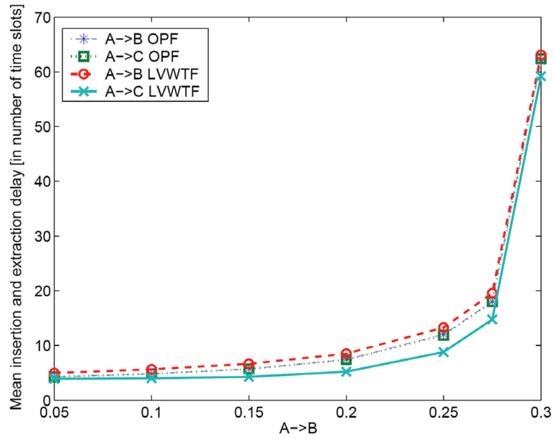

Fig. 4. Performance of scheduling when $A \rightarrow B \leq 0.3$ and $A \rightarrow C = 0.1$.

Finally, it may also happen that the dimensioning process has underestimated some actual flow rates. We argue here that this is not an issue that can be suitably handled by the scheduling policy because the dimensioning itself may not be stable anymore! Supplementary policies, such as e.g. policing or conformance control mechanisms, are requested to deal with these cases, and to limit the actual flow rates to their agreed limits.

B. Performance for a multi-wavelength network

In order to illustrate this more general case, we consider a 6-node 2-wavelength ring where all stations can listen on both wavelengths. We label the ring nodes with A, B, C, D, E and F, in traffic direction. The traffic matrix is complete, symmetric and uniform of rate *x* with one exception: node A is the only node sending non-symmetric traffic. Node A sends traffic of load *x* to stations C, D, E and F, and traffic of load 2*x* to station B.

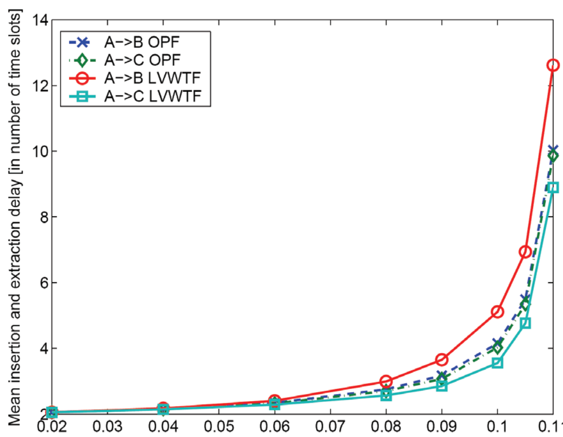

Fig. 5. LVWTF versus OPF in 2-wavelength receiver ring: $A \rightarrow B = 2 \cdot (A \rightarrow C)$.

In Fig. 5, the expected waiting times for flows AB and AC are compared by simulation in cases of LVWTF and OPF policy, in function of traffic rate $A \rightarrow C$. The differences in the latency between OPF and LVWTF remain limited in this scenario, which confirms that LVWTF is a good candidate for approximating OPF.

V. Conclusion

In this paper we have addressed the issue of selecting scheduling policies for a slotted WDM optical packet ring. We have first stated that, although desirable in terms of delay performance, FIFO scheduling is not appropriate as it presents head-of-the-line blocking. After pointing out the necessity of selecting a MaxWeight scheduling policy, we have compared the performance delivered by LQF and LVWTF to the one delivered by OPF which closely emulates FIFO.

We have shown that OPF is far more complex to implement than LQF and LVWTF, and that LVWTF, which is as simple as LQF, provides a queueing performance close to the one delivered by OPF. Actually, the differences between OPF, LQF and LVWTF can be shown to decrease when the number of wavelengths in the ring increases. This is due to the multiplexing gain brought by WDM as shown in a previous paper [10].

Acknowledgment

The work described in this paper was carried out with the support of the French "Agence Nationale de la Recherche" in the framework of the ECOFRAME project (grant number 2006 TCOM-002-06), and of the BONE Network of Excellence ("Building the Future Optical Network in Europe"), funded by the European Commission through the 7th ICT-Framework Programme.

References

[1] I. White *et al.*, "A summary of the HORNET project: a next-generation metropolitan area network," *IEEE Journal on Selected Areas in Communications*, vol. 21, pp. 1478–1494, Nov. 2003.

[2] A. Carena *et al.*, "RingO: an experimental WDM optical packet network for metro applications," *IEEE Journal on Selected Areas in Communications*, vol. 22, pp. 1561–1571, Oct. 2004.

[3] D. Dey *et al.*, "FLAMINGO: A Packet-switched IP-over-WDM All optical MAN," *ECOC 2001*.

[4] A. Stavdas *et al.*, "IST-DAVID: concept presentation and physical layer modeling of the metropolitan area network," *Journal of Lightwave Technology*, vol. 21, pp. 372–383, Feb. 2003.

[5] J. Cai, A. Fumagalli, and I. Chlamtac, "The multitoken interarrival time (MTIT) access protocol for supporting variable size packets over WDM ring network," *IEEE Journal on Selected Areas in Communications*, vol. 18, pp. 2094–2104, Oct 2000.

[6] G. Hu *et al.*, "Performance of MAC layer and fairness protocol for the dual bus optical ring network (DBORN)," *Conference on Optical Network Design and Modeling*, pp. 467–476, 7-9, 2005.

[7] M. Herzog *et al.*, "RINGOSTAR: an evolutionary performance enhancing WDM upgrade of IEEE 802.17 resilient packet ring," *IEEE Communications Magazine*, vol. 44, pp. 8–14, Feb. 2006.

[8] J. A. Lazaro *et al.*, "Remotely amplified SARDANA: Single-fibre-tree advanced ring-based dense access network architecture," *ECOC 2006*.

[9] B. Uscumlic, A. Gravey, P. Gravey, and M. Morvan, "Impact of Peer-to-Peer Traffic on the Efficiency of Optical Packet Rings," *WOBS 2008*.

[10] B. Uscumlic, A. Gravey, P. Gravey, and I. Cerutti, "Traffic Grooming in WDM Optical Packet Rings," *ITC 21*, 2009.

[11] B. Uscumlic, A. Gravey, I. Cerutti, P. Gravey, and M. Morvan, "The Impact of Network Design on Packet Scheduling in Slotted WDM Packet Rings," *Photonics in Switching*, 2009.

[12] A. L. Stolyar, "MaxWeight Scheduling in a Generalized Switch: State Space Collapse and Workload Minimization in Heavy Traffic," *Annals of App. Prob.*, vol. 14, no. 1, pp. 1–53, 2004.

[13] L. Tassiulas and A. Ephremides, "Dynamic server allocation to parallel queues with randomly varying connectivity," *IEEE Transactions on Information Theory*, vol. 39, March 1993.

[14] B. Uscumlic, A. Gravey, I. Cerutti, P. Gravey, and M. Morvan, "Stable dimensioning of optical packet rings," *unpublished*.

[15] T. Eido *et al.*, "Packet Filling Optimization in Multiservice Slotted Optical Packet Switching MAN Networks," *AICT*, 2008.

[16] N. McKeown *et al.*, "Achieving 100% Throughput in an Input-Queued Switch," *Proc. IEEE INFOCOM*, 1996.