Resource Management in Optical Burst Switched Networks: Performance Evaluation of a European Network *

Maurizio Casoni[a] and Maria Luisa Merani[a]

[a]Department of Information Engineering - University of Modena and Reggio Emilia
via Vignolese, 905 - 41100 Modena (Italy)

ABSTRACT

This work investigates the edge-to-edge performance of an Optical Burst Switched network employing a reservation mechanism known as JET, Just Enough Time. Two main components are studied and analyzed, the edge router with the burst assembly function and the core network with routing and forwarding. We assess the performance, in terms of burst blocking probability and delay, experienced by three service classes on a proposed European network.

Multiple traffic source descriptors are considered. Different traffic settings, resource management and routing options are investigated with the aim to meet diverse QoS requirements.

Keywords: Optical Burst Switching, Resource Management, Differentiated QoS, QoS Routing, Self-Similar Traffic

1. INTRODUCTION

Optical networking has been boosted by the advances in coherent optical transmission, which led to Dense Wavelength Division Multiplexing (DWDM) systems able to accommodate up to hundreds of wavelengths per fiber, and in integrated optics, for passive and active optical components design. The ultimate goal is the development of a fully optical Internet, where signals carried within the network never leave the optical domain. A first important step in this direction is to have optical networks transparent at least for data, with the control part converted and processed in electronics. In addition, as current and future applications require different network performance in terms of, e.g., loss, delay, jitter, network design becomes more complex.

In this paper the focus is on the performance of an optical switching technology which offers a dynamic mechanism to set up high capacity end-to-end optical data connections in DWDM networks: the Optical Burst Switching (OBS) solution.[1, 2] In OBS networks, data never leave the optical domain: for each data burst assembled at the network edge, a reservation request is sent as a separate control packet, well in advance. Optical burst switching can therefore be seen as a middle term solution towards all optical packet switching,[3] fulfilling the goal to improve wavelength utilization and sharing by introducing a dynamic wavelength management. The key idea behind OBS is to dynamically set up a wavelength path whenever a *large* data flow is identified and needs to traverse the network: a separate control packet therefore precedes each burst by a basic offset time, carrying relevant forwarding information. It is this out-of-band signaling solution that allows to perform resource allocation within the electronic domain, as a proper offset between the control packet and the burst gives the traversed nodes enough time to process the control information and to set-up the optical cross-connects (OXCs).

In recent years several papers have contributed to the definition and study of the OBS paradigm. To name a few,[1] and[5] provide a technological engineering framework for the OBS architecture;[4] and[6] address the issue of guaranteeing basic quality of service in OBS networks and also propose the use of LOBS, Label Optical Burst Switching, as the control solution under the IP MPLS, MultiProtocol Label Switching, paradigm.

---

*This work has been partially supported by MIUR within the framework of the national project "INTREPIDO: Traffic Engineering and Protection for IP over DWDM".

Further author information: (Send correspondence to Maurizio Casoni)
Maurizio Casoni: E-mail: casoni.maurizio@unimore.it, Telephone: +39 059 205 6167

For those reservation strategies which do not aim at differentiating service levels in OBS nodes and networks, the offset duration is exclusively determined by the different edge-to-edge delays faced by the control packet and data. For the reservation mechanism known as JET, the OXC allocates a specific input-output path to a burst, only for the burst duration; furthermore, the offset might also account for an additional contribution which allows to differentiate the performance of various classes of service and grant them different priorities. The principle is that the longer the extra-offset, the higher the priority, as the corresponding flow can reserve very in advance the required resource. Hence, JET can also enforce QoS differentiation directly at the optical layer.[4]

In OBS networks, nodes can be classified as either edge or core routers. The main task of edge nodes is the burst assemble function: as they represent the border between "traditional" electrical LAN/MAN IP networks and a high speed optical transport network,[3] they must collect incoming IP datagrams and assembly them into bursts according to suitable algorithms. These must take into account QoS requirements, final destinations and technological constraints. Many open issues are currently investigated for proper settings, such as assembly time and extra-offset .[5, 7–10] A first contribution this paper provides is to investigate a specific burst assembly mechanism, in order to get some hints on the additional delay the assembler introduces and on the burst length distributions.

Core routers, on the other hand, deal with data bursts and the related control packets; they have to set-up on the fly internal optical paths for switching bursts and take them hop-by-hop closer to their final destination. Core routers keep data bursts in the optical domain, whereas control packets are O-E converted for processing. The main issues such routers have to deal with, when data bursts are considered, are output contention resolution, QoS support in burst forwarding and QoS routing. To this regard, a second contribution of this paper is to propose and investigate a simple but feasible OBS node equipped with input fiber delay lines and a limited set of wavelength converters: the proposal is to allow only delay-sensitive bursts to employ such converters, in order to grant them a better service. Further, edge-to-edge performance is evaluated when such node is employed as a core router in a European OBS network. Dijkstra routing algorithm is first adopted, then complemented by deflection routing, allowed for best-effort burst classes.

Many aspects are thus investigated, with the aim to get an insight on the setting of edge and core routers for the support of different QoS classes. Concerning the structure of the paper, Section 2 explains the edge nodes and details the burst assembly mechanism, also providing a description of the incoming traffic pattern. In Section 3 a single OBS core node is described and a network with this node as a building block is investigated. Section 4 reports some numerical results with reference to different operating conditions, while the main outcomes that have been derived are summarized in Section 5.

2. EDGE ROUTERS AND BURST ASSEMBLY

This work assumes that OBS nodes support JET and that traffic is mapped into three QoS classes. Each of them is characterized by a different statistical traffic description and features different QoS requirements, i.e., specifying an upper bound of burst blocking probability and edge-to-edge delay.

In this work, incoming traffic from legacy LANs is supposed to exclusively consist of IP datagrams: according to,[11] medium and large-size datagrams are 576 and 1500 byte long; a third IP datagram size, equal to 200 bytes, is introduced to account for Voice over IP applications (20 bytes of IP header, 8 bytes of UDP header, 12 bytes of RTP header and 160 bytes of G.711 payload). Furthermore, small packets, $40 - 44$ byte long, that represent a relevant percentage of Internet backbone traffic, are properly assembled into bursts as well. As noted in the Introduction, the burst assembler investigated in this paper is reported in Figure 1. Incoming datagrams are grouped according to their destination and three classes of bursts, labelled class-1, class-2 and class-3, are defined: class 1 is created with time-sensitive data of VoIP sessions and with data that allow the management of closed-loop flow control, i.e., ACKs of end-to-end TCP flow control. In the burst assembly process class 1 bursts are attributed the highest priority and thus given the longest extra-offset time. Class 2 and 3 bursts are made of 576 and 1500 bytes datagrams respectively, are supposed to be loss-sensitive and no extra-offset is given to them.

The employed burst assembly algorithm is of the T_{max}-L_{min} type: depending on the nature of the information flow, whether time sensitive or loss sensitive, two different goals are therefore pursued. For the former, as it is the edge-to-edge delay to be bounded, a maximum delay T_{max} can be tolerated in the assembly phase, but after that the burst must be transmitted. For the latter delay is not a constraint: the main goal is to achieve the highest efficiency by transmitting bursts only *after* they have reached a minimum length L_{min}. This also implies to keep the burst length under control, avoiding too long bursts which may adversely impact network performance. Thus, bursts must be created "adequately" long and, in addition, a maximum assembly delay must be considered in order not to penalize end-to-end throughputs. Next, a scheduler S takes the bursts from the burst assembly blocks as soon as they are ready and send them to the optical output interface for transmission on any free available wavelength. Two more levels of priority will be successively introduced for class 2 and 3 bursts, enforced through a suitable policy of employing the limited set of wavelength converters.

In Section 4, some numerical evaluations in terms of burst length will be reported showing the traffic pattern generated by the assembly mechanism previously described.

3. CORE ROUTERS AND NETWORK

3.1. Core Node and Traffic Description

The core router investigated in this paper is equipped with $M \times M$ optical interfaces capable of supporting N wavelengths each. Over each wavelength optical transmission devices allow for transmissions at 10 Gbit/s. We suppose that some wavelength converters are available and that the single optical node is bufferless, i.e., no fiber delay lines (FDLs) are available in order to resolve contention for an output fiber, output wavelength (Figure 2). When the JET reservation scheme is employed, the setup of the basic offset time has to be carefully done in order to allow for the data burst to go through the intermediate nodes and to reach successfully the destination node without overtaking the associated control packet. Let T_{p_i} and T_{p_d} be the processing time of a control packet at an intermediate and destination node, respectively. Thus, for N_{hops} traversed nodes, the basic offset time Δ should at least be equal to

$$\Delta = \sum_{k=1}^{N_{hops}} T_{p_i}^{k} + T_{p_d} \, . \tag{1}$$

Generally, this implies determining in advance the path within the core network, as it is necessary to know the number N_{hops} of hops, or equivalently how many intermediate nodes, the burst will cross. In a datagram network such as the Internet, this information is usually not so easy to derive, and it would be desirable to make Δ independent of the number of hops. A bufferless node is assumed, however it is reasonable to consider the use of a set of input FDL,[5] whose tasks is to re-align the data burst and its control packet at the I-th intermediate nodes so as to guarantee an offset time at least equal to T_{p_i}. This makes, as desired, Δ independent of the path and allows to set Δ to $max\{T_{p_d}^{j}\}$, for all j destinations.

Regarding different priority levels, JET offers a way to reduce the blocking of higher priority users granting them a longer extra-offset time. If the extra-offset time is properly set, highest priority bursts are almost exclusively blocked by bursts belonging to the same class; the blocking they experience is therefore likely to be reduced to its minimum.

When the LOBS approach is employed, the label carried by the control packet associated to the incoming data flow simply indicates the next hop to which data has to be forwarded. Equivalently, bursts are switched to the desired output interface – fiber – regardless of the output wavelength. It is worth underlining that wavelength converters are necessary in order to allow an incoming data flow on any input wavelength to be switched to any wavelength of a generic outgoing fiber. A preliminary study of single OBS node in presence of ON/OFF traffic on each input wavelength and full wavelength conversion has been previously performed.[12] In this paper a limited set of such converters is assumed: moreover, only class 1 and 2 bursts are allowed to employ these devices. Low priority class 3 bursts, on the other hand, cannot exploit them and can be forwarded to the desired output link on the same wavelength only. Of course, this may degrade their performance.

On each input wavelength let λ represent the mean overall arrival rate of bursts. Given that $\overline{x}_1$, $\overline{x}_2$ and $\overline{x}_3$ are the mean duration times of the ON periods due to bursts of short, medium and long-sized datagrams, respectively, on each wavelength the offered load ρ is the sum of three distinct contributions,

$$\rho = \sum_{i=1}^{3} \rho_i = \sum_{i=1}^{3} p_i \lambda \overline{x}_i \, , \tag{2}$$

where p_1, p_2 and p_3 represent the occurrence probabilities of bursts of the three classes.

Aiming to model self-similar traffic,[13] Pareto-distributed ON periods are considered. Indicating by X_{on_i} the random variable which represents the duration on the ON period of class i bursts, its cumulative distribution function is

$$F(x) = P[X_{on_i} \leq x] = 1 - \left(\frac{k_{on_i}}{x} \right)^{\alpha_{on_i}} , 0 < k_{on_i} \leq x \, , \tag{3}$$

where the constant k_{on_i} is the smallest possible value of the random variable X_{on_i}, expressed as multiple of a basic unit (here datagram size of class i); $0 < \alpha_{on_i} < 2$ in order to have a heavy-tailed distribution. Note that under the hypothesis $0 < \alpha_{on_i} < 2$, X_{on_i} has infinite variance. Equivalently, one can affirm that the probability of very large random sampling of X_{on_i} is not negligible. In computer networks, usually $1 < \alpha_{on_i} < 2$ and the corresponding mean value is

$$E[X_{on_i}] = \frac{\alpha_{on_i} k_{on_i}}{\alpha_{on_i} - 1} \, . \tag{4}$$

A measure of the degree of self-similarity is the Hurst parameter, $H(\alpha) = \frac{3-\alpha}{2}$, a dimensionless quantity which ranges between 0.5 (no self-similarity, $\alpha = 2$) and 1 ($\alpha = 1$). OFF periods are exponentially distributed.

3.2. Routing and Network Topology

Regarding the network as a whole, information flows are made of bursts routed within the network by means of the Dijkstra algorithm. Since bursts have been created according to three classes of service, class 1 carrying time-sensitive data, class 2 and 3 loss-sensitive data, routing is modified in order to better meet the performance required by each class and a simple QoS routing criterion is introduced: class 2 and 3 bursts are allowed to be deflected and thus re-routed in case of unavailability of a wavelength in the desired output fiber. This may lead to a variable edge-to-edge delay which needs to be evaluated; however, this solution is better than increasing the loss of this type of bursts. It is worth noting that variable delays can be managed by the node architecture assumed (Figure 2), employing input FDLs.

In summary, class 1 bursts are given the highest priority through an additional extra-offset and the use of wavelength converters; class 2 bursts have medium priority by using the converters; class 3 bursts have low priority since they have less offset and cannot exploit wavelength conversion; on the other hand, class 2 and 3 can use alternative sub-optimal variable delay paths.

In this paper the network considered covers most European countries (Figure 3). Each node operates as a core router and, in addition, nodes A (London), B (Oslo) and C (Stockolm) are sources of information flows, whereas nodes O (Madrid), P (Rome) and Q (Athens) are the possible destinations; moreover, no flow is supposed to enter or leave the network at intermediate steps.

Next section will show the performance, in terms of burst blocking probability for a single node, as well as the overall edge-to-edge burst blocking probability and the delay for the three classes of bursts in the OBS network described above, when each node implements the JET reservation mechanism with a limited set of wavelength converters available, under different mixtures of traffic.

4. NUMERICAL RESULTS

In this Section we present some results referring to both the burst assembly function in edge nodes and to the OBS network reported in Figure 3. The performance of this network, including the edge nodes, has been studied through an *ad-hoc* event-driven C++ object oriented simulator; for the results regarding burst loss probability, a 95% confidence interval has been determined and the t-Student distribution has been employed to assess their reliability. Burst blocking probability of class i is here determined as the ratio between the number of class i burst lost and the total offered burst of class i.

Edge routers perform the burst assembly function: on one side they have 10 GEthernet cards and on the other they have M outgoing fibers, with N wavelengths per fiber. The main building blocks are reported in Figure 1. Regarding the edge router performance and the assembly mechanism, input traffic exhibits the following characteristics: $p_1 = 0.6$, given by 0.3 of VoIP traffic and 0.3 of small sized datagrams, $p_2 = 0.2$ and $p_3 = 0.2$, with datagram lengths equal to 200 and 64 bytes for class 1, 576 and 1500 bytes for class 2 and 3.

Figure 4 shows the assembly time for class 3 (loss-sensitive) bursts as a function of the offered traffic ρ for different values of L_{min}, ranging from 640 kbytes to 10 Mbytes. T_{max} is set to 20 msec. Information contained in this kind of bursts is very likely to be controlled by a reliable end-to-end flow control mechanism, e.g. TCP, whose performance also depends on the round trip time, in turn influenced by the delay added by burst assembly. This Figure shows that at light loads, e.g. $0.2 - 0.3$, different settings of L_{min} lead to delays ranging from 1 to 17 msec, which may impact end-to-end performance in a markedly different way. In other words, it is possible to select L_{min} as a function of traffic load to incur given assembly delay. Furthermore, L_{min} allows to keep under control the size of the bursts injected into the network, thus affecting the end-to-end performance, measured by the burst blocking probability. Investigating the impact of the assembly algorithm on the performance of the OBS node and of the network is outside the scope of this paper. Instead, we investigate their behavior when the input traffic follows a heavy-tailed distribution, so as to presumably evaluate a "worst case" performance.

Let us first consider a single optical node with $M = 4$ incoming and $M = 4$ outgoing fibers, with $N = 8$ wavelengths per fiber. Each of the outgoing fibers is chosen with equal probability $p_F = 1/M = 1/4$. The three examined service classes generate bursts composed of datagrams of length 200, 576 and 1500 bytes transmitted at a transmission rate equal to $B_r = 10$ Gbit/s per wavelength. It is assumed $\alpha_{on_i} = \alpha_{on}$, the same for the three classes. Input traffic, generated as detailed in the previous Section, has been tested to verify if it exhibits a heavy tailed distribution. As reported in[13] chapter 3, the simplest way to assess the presence of heavy tailed distribution is to plot the complementary distribution function (ccdf) on log-log axes. Figure 5 shows the ccdf of sizes of class 3 bursts for three values of k_{on_3}, $k_{on_3} = 1, 5, 15$. In all cases this Figure shows a straight line on log-log axes with a slope $-\alpha_{on}$, being $\alpha_{on} = 1.2$ the adopted value. Similar plots can be drawn for the other two classes, differing each other for a scaling factor. Here the following values for p_1, p_2 and p_3 are considered, $(0.2, 0.5, 0.3)$; for a given input load ρ and a specific (p_1, p_2, p_3) set, the actual value of λ is derived from eq.(2). In the following, $\forall i k_{on_i} = 1$ is assumed.

As our approach for service differentiation consists of jointly managing extra-offset values and wavelength conversion option, it is assumed that classes 1 and 2 only are allowed to employ wavelength converters, thus penalizing class 3; in addition, class 1 is also assigned an extra-offset value, as it is supposed to be the high priority class serving time-sensitive applications. Each core router is therefore equipped with a set of WC ($WC < N \times M$) wavelength converters available to class 1 and class 2 bursts only. Figure 6 shows the burst blocking probability of class 1, 2, 3 as a function of the offered load with a set of $WC = 20$ wavelength converters. This value has been verified to represent the optimum, meaning that no additional benefits are obtained by considering a larger set for this node setting. An extra-offset equal to $3.6\mu s$ (i.e., the transmission time of 4500 bytes at 10 Gbit/s) is assigned to class 1. It is worth noting the remarkable differentiation our approach achieves: even if the low priority "best effort" class 3 almost always gets a 50% burst blocking, class 1 is always below 10^{-3} and if the load is limited to 0.4 also class 2 bursts experience an excellent 2% blocking.

Consider now the network reported in Figure 3. Information flows are made of bursts routed within the network by means of the Dijkstra algorithm and flowing along the North-South direction. Since bursts have been created according to three classes of service, class 1 carrying time-sensitive data, class 2 and 3 loss-sensitive data, routing is modified in order to better meet the performance required by each class and a simple QoS routing criterion is conceived: class 2 and 3 bursts are allowed to be deflected in case of unavailability of a wavelength on the desired output fiber. A generic node within the network has 4 incoming and 4 outgoing fibers, with 8 wavelengths per fiber. Given the particular topology, nodes connected to two nodes of the next step symmetrically assign two fibers per each connection. A-B-C (O-P-Q) nodes on one side have 10 GE cards and on the other $M = 4$ outgoing (incoming) fibers, with $N = 8$ wavelengths per fiber. In order to differentiate service among classes, class 1 bursts are assigned an extra-offset equal to $5 \times \overline{x}_3 = 6$ μs: this value allows for a class isolation degree greater than 99% when burst lengths are exponentially distributed.[4] Although the traffic pattern considered in this study is different, we assume this same value. Class 2 and 3 bursts employ no extra-offset. Δ is assumed to be 0.5μs. As for the single node, the following values for p_1, p_2 and p_3 are considered, $(0.5, 0.2, 0.3)$; moreover, incoming traffic is supposed to have ON/OFF periods with OFF exponentially distributed and ON following the Pareto distribution with the same parameters as above. Furthermore, only class 1 and 2 bursts exploit the set of 20 wavelength converters; however, class 2 and 3 can be re-routed in case of unavailability of wavelengths on the outgoing fibre along the least-cost path.

Figure 7 shows the total burst blocking probability for the three burst classes having node B as source and node P as destination. In order to have loss values in the range of 10^{-3} for class 1, 1% for class 2 and 50% for class 3, the overall load ρ must be less than 0.4.

Figure 8 shows the edge-to-edge delay, measured in hop counts, for the three classes having node B as source and node P as destination. While class 1 bursts experience a fixed delay, and this is exactly what is required by time-sensitive data, class 2 and 3 bursts suffer a variable delay, due to the ”on-the-fly” re-routing after a lost contention during forwarding inside a core router. This phenomenon is more evident for class 3 bursts, as they are assigned the lowest priority, and more often ask for alternative paths.

5. CONCLUSIONS

This work has evaluated the performance experienced by three classes of service in an OBS network implementing the JET reservation mechanism. Some of the most relevant issues encountered in an OBS network have been presented and investigated: edge node with the burst assembly function, single core router, a sample European network with OBS nodes adopting a simple QoS routing.

The single node and the end-to-end burst blocking probabilities, the edge-to-edge delay have been studied, as well as the burst assembly mechanism required at the edge nodes when the $T_{max} - L_{min}$ method is employed. It has been shown that a careful tuning of the burst assembly algorithm is needed in order to be efficient and not to penalize loss sensitive data. Furthermore, a simple QoS routing policy has been proposed, to keep acceptable also the performance of low priority bursts by means of re-routing.

The proposed approach for service differentiation consisting of a proper management of extra-offset and wavelength converters, also exploiting class-based routing, allows for different performance levels and provides useful insights for traffic and network engineering.

Our current efforts are focused on a complete integration of all network elements with the goal of developing a flexible and dynamic simulation tool and on the development of analytical models for the OBS node and network.

REFERENCES

1. J. S. Turner, “Terabit Burst Switching,” *Journal of High Speed Networks*, No.8, pp.3-16, 1999.
2. C. Qiao, M. Yoo, “Optical Burst Switching (OBS) - a New Paradigm for an Optical Internet,” *Journal of High Speed Networks*, No.8, pp.69-84, 1999.
3. F.Callegati, M.Casoni, C.Raffaelli, ”Packet optical networks for high-speed TCP-IP backbones”, *IEEE Communication Magazine*, Volume 37, Issue 1, January 1999, pp. 124-129.
4. M. Yoo, C. Qiao, S. Dixit, “QoS Performance of Optical Burst Switching in IP-Over-WDM Networks,” *IEEE Journal on Selected Areas in Communications*, Vol.18, No.10, pp.2062-2071, October 2000.
5. Y. Xiong, M. Vandenhoute, H.C. Cankaya, ”Control Architecture in Optical Burst-Switched WDM Networks”, *IEEE Journal on Selected Areas in Communications*, Vol.18, No.10, October 2000, pp.1838-1851.

6. C. Qiao, “Labeled Optical Burst Switching for IP over WDM integration,” *IEEE Communication Magazine*, September 2000, pp.104-114.
7. V.M. Vokkarane, Q. Zhang, J.P. Jue, B. Chen, ”Generalized Burst Assembly and Scheduling Techniques for QoS Optical Burst-Switched Networks”, Proc. of Globecom 2002, November 2002, Taipei (Taiwan).
8. A. GE, F.Callegati, L.S. Tamil, ”On Optical Burst Switching and Self-similar Traffic”, *IEEE Communications Letters*, Vol. 4, No. 3, March 2000, pp.98-100.
9. X. Yu, Y. Chen, C. Qiao, ”A Study of Traffic Statistics of Assembled Burst Traffic in Optical Burst Switched Networks”, Opticomm 2002, Proc. of SPIE Vol. 4874 (2002), pp.149-159
10. X. Cao, J. Li, Y. Chen, C. Qiao, ”Assembling TCP/IP Packets in Optical Burst Switched Networks”, Proc. of Globecom 2002, November 2002, Taipei (Taiwan).
11. K. Thompson, G. J. Miller, R. Wilder, “Wide-Area Internet Traffic Patterns and Characteristics,” *IEEE Network*, November/December 1997, pp.10-23.
12. A. Brandani, M. Casoni, M.L. Merani, ”MPλS vs. LOBS in an Optical Burst Switching Node: Impact of Long-Range Dependant Traffic on Multiclass Services”, Proc. of IEEE ICOCN 2002, November 2002, Singapore, pp.186-189.
13. K. Park, W. Willinger, *Self-Similar Network Traffic and Performance Evaluation.* J.Wiley & Sons: New York, 2000.

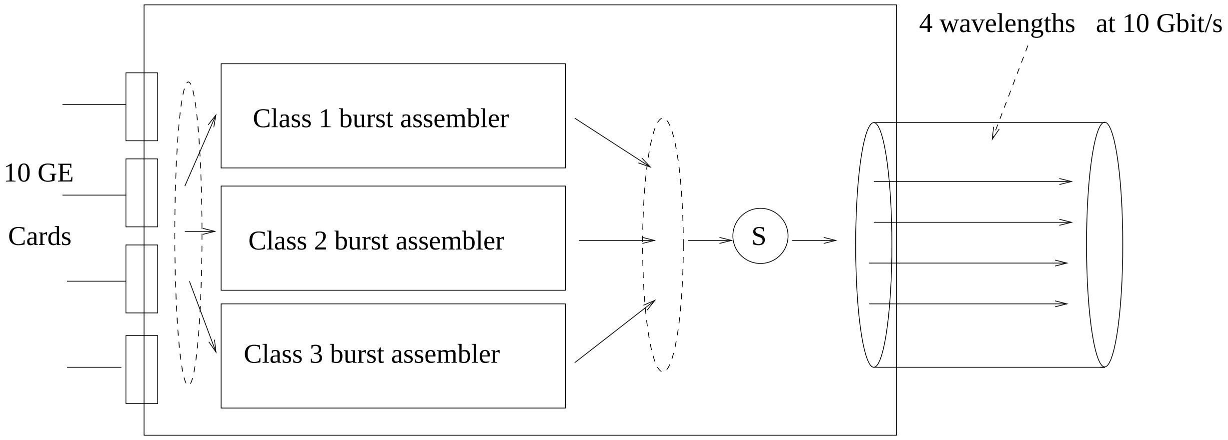

Figure 1: The burst assembler in edge routers.

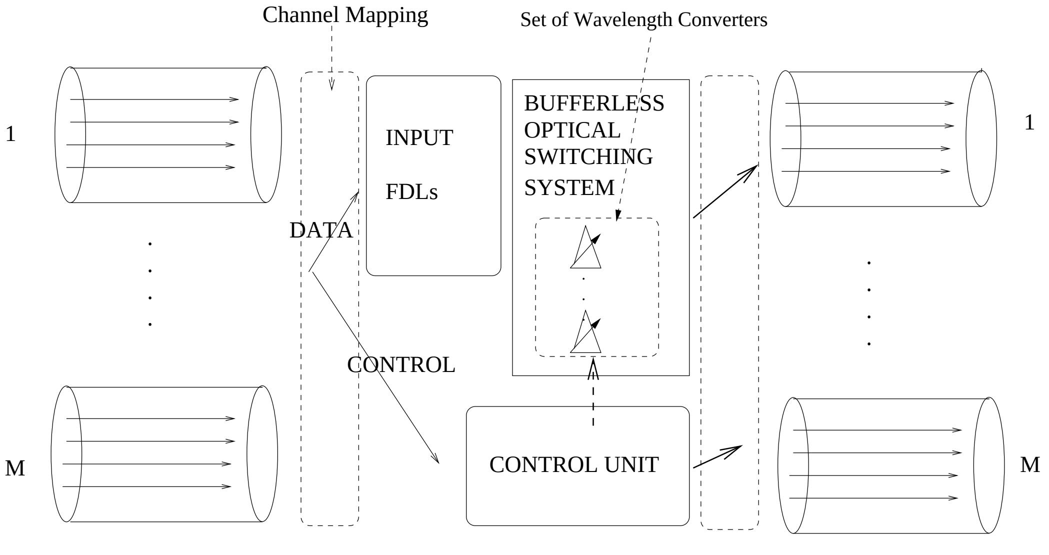

Figure 2: General core router architecture.

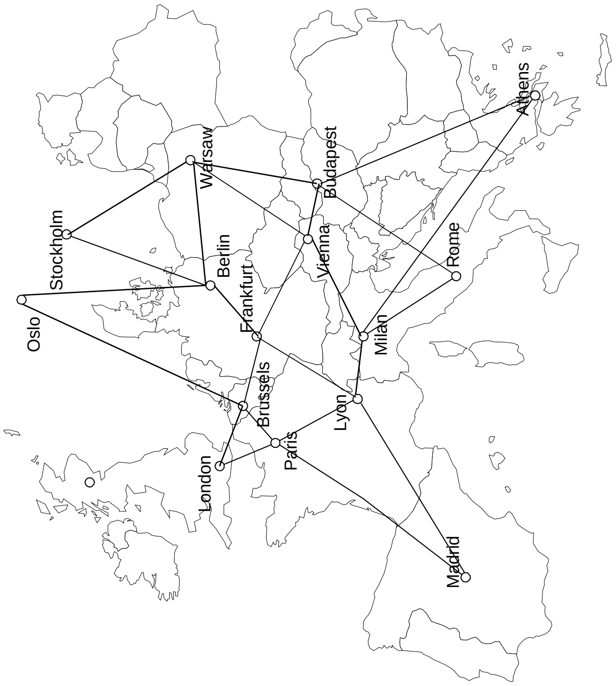

Figure 3: The reference network.

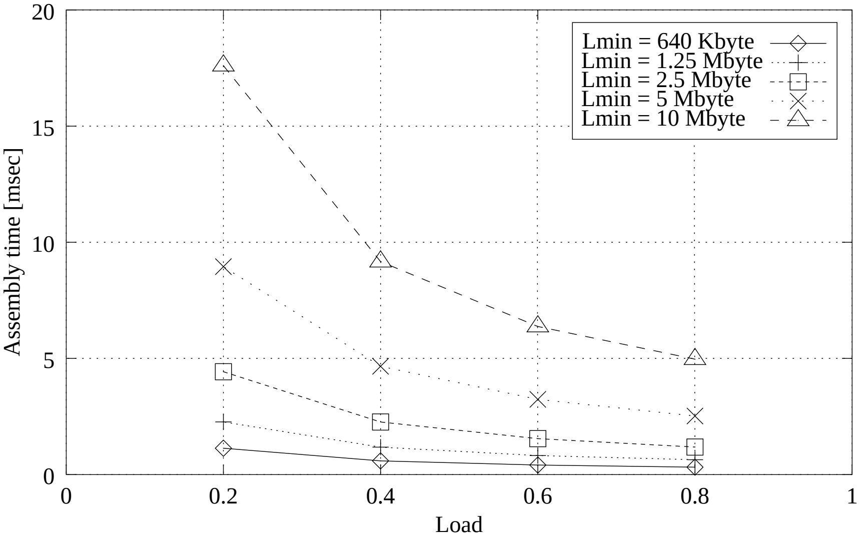

Figure 4: Class 3 assembly period as a function of the offered load for different values of the minimum burst length.

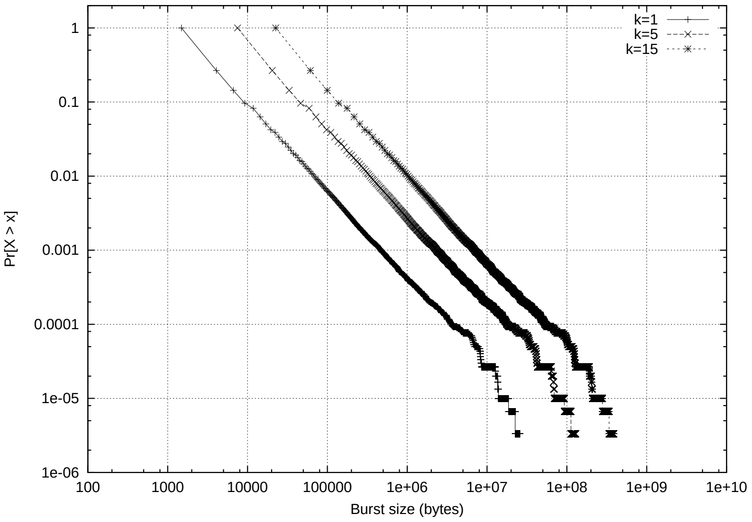

Figure 5: Log-log ccdf of burst size for class 3 and $K_{on} = 1, 5, 15$.

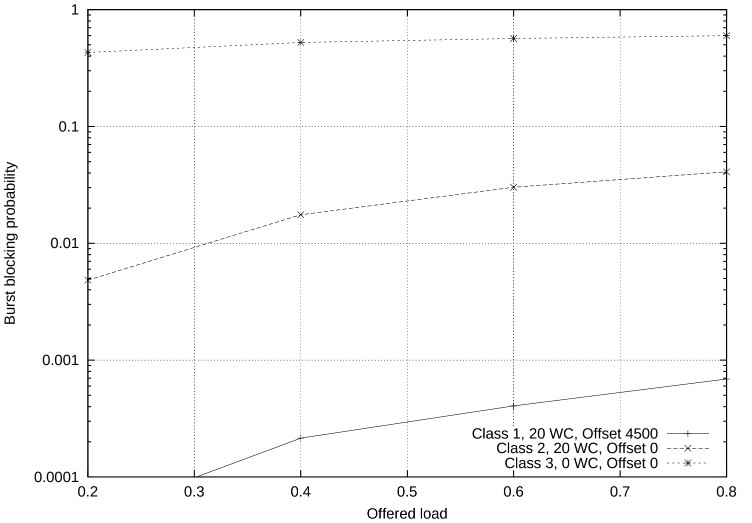

Figure 6. Burst blocking probability of class 1, 2, 3 vs. offered load with a set of 20 wavelength converters. Extra-offset employed for class 1 only.

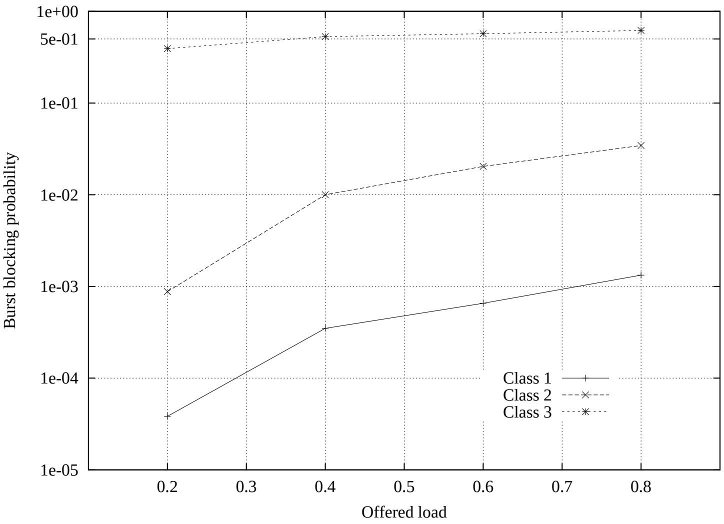

Figure 7. Total End-to-end burst blocking probability as a function of the offered load for the three burst classes, with reference to the (B,P) pair

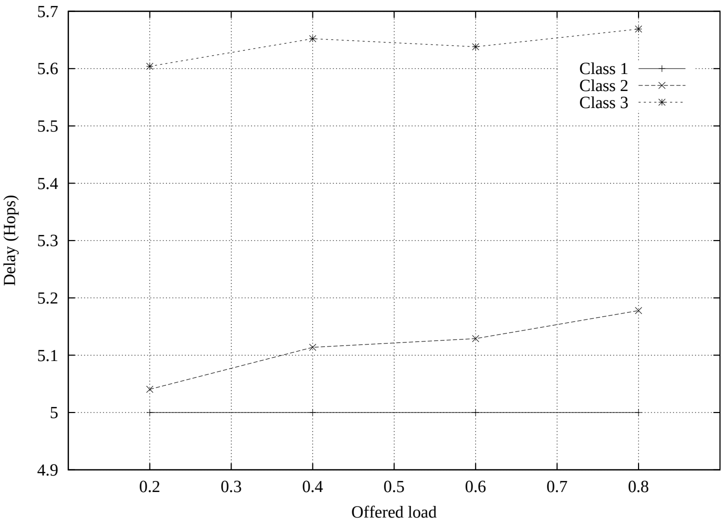

Figure 8. Edge-to-edge delay (hop counts) for the three burst classes as a function of the offered load, with reference to the (B,P) pair.