A Translucent OBS Node Architecture to Improve Traffic Emission and Loss Probability

Thomas Coutelen
ECE, Concordia University
Montreal, Qc, H3G 1M8, Canada
Email: t_coutel@ece.concordia.ca

Brigitte Jaumard
CIISE, Concordia University
Montreal, Qc, H3G 1M8, Canada
Email: bjaumard@ciise.concordia.ca

Gérard Hébuterne
Télécom-SudParis, UMR CNRS Samovar
Evry, France
Email:gerard.hebuterne@it-sudparis.eu

Abstract—**All-optical circuit switching prevents contention by forbidding multiplexing beyond the wavelength granularity. The drawbacks of such a coarse granularity can be reduced thanks to translucent architectures with MSPPs (MultiService Provisioning Platforms). Another switching paradigm of interest is offered by the Optical Burst Switching (OBS) that manages finer granularity transfers but that encounters contentions, potentially entailing high loss rates.**

In this paper, within the context of core networks, we describe a translucent OBS architecture that performs intermediate re-aggregation and allows the reduction of the loss rate by interceding at two levels: Firstly, it provides sparse electrical buffering accessibility and, secondly, it uses aggregation grooming in order to reduce the aggregation delay and the traffic contention probability. A careful analysis of the traffic behavior in core networks leads us to propose an accurate traffic model, called LCH$^+$, i.e., an enhanced Lost Call Held (LCH) model. Data plane transparency is then exploited to modulate the traffic at its emission. Extensive experiments show that re-aggregation helps reducing the loss probability thanks to electrical buffers whereas the aggregation grooming mitigates the drawbacks induced by electrical buffering on the delay and on the network cost. In addition, we show that the proposed translucent architecture can be highly valuable in clustered large core networks.

I. Introduction

Optical Burst Switching (OBS, [1]) has been proposed in the late 90s to address the coarse granularity issue of OCS (Optical Circuit Switching) that can severely undermine the transport capacity utilization under bursty or dynamic traffic. An OBS network manages bursts (i.e., large aggregate of IP packets) instead of circuits and, consequently, opens the possibility of statistical multiplexing. However, the lack of optical memory dictates a transparent data plane. This is made possible by the use of in advance signaling protocols such as the JET (Just Enough Time) one.

The control plane load is reduced by the aggregation process: Several IP packets are aggregated into a burst and therefore signaled by a single header. The basic aggregation algorithm (see [2]) has been adapted to meet specific requirements related to, i.e., delay [3] or Quality of Service (QoS) [4]. Transfer latency is reduced by the use of a one way reservation. Unfortunately, contentions can occur and their avoidance or resolution is the key challenge in OBS networks and the most critical performance criterion. The contention resolution mechanisms operate either in time, space or spectral domains, and, even if domains are combined, they cannot guarantee lossless transfers [5]. Pro-active mechanisms aiming at reducing the contention rate are highly beneficial [6], [7] but remain insufficient.

The aggregation process is crucial in an OBS architecture and must be carefully designed as it impacts the OBS performances. For example, burst size directly strikes the aggregation delay, the FDL (Fiber Delay Line) efficiency [8], the signaling overhead, and the traffic shaping. Atop these issues, the definition of the aggregation queues is important. In [9], authors show that distributing the incoming traffic among several aggregation queues improves the goodput of TCP sources by reducing the impact of the synchronized loss. All those considerations are relevant, but their importance depends on the equipment and protocols. In this paper, the design of the aggregation process will be driven by its impact on the traffic profile in a core network.

Because of the data plane transparency, the traffic profile is not affected by switching operations and is thus completely defined at its emission. A deep analysis of the traffic in a core network, from the data plane transparency up to the burst arrival process in the core nodes has led us to the proposal of the LCH$^+$ (Lost Call Held) model that accurately includes OBS specificities. Experiments put forward a reduction of the number of overlapping bursts in the network.

A study of the aggregation and of the MAC processes show that the burst sizes and the number of aggregation queues are parameters that strongly impact the overlapping occurrence in a core network. Grooming must be considered to sidestep the stringent definition of aggregation queues imposed by the topology. In the context of OBS, it has first been explored in [10] where "core grooming" is achieved with FDLs: Successive bursts with the same destinations are moved closer so that they can be handled as a single longer burst. The work of [10] lingers the negative impact of small bursts, but does not involve an aggregation process per se. In [11], small bursts are avoided at the aggregation by padding them with traffic from other connections. The electrical processing, mandatory to split bursts at some node is avoided in [12] thanks to optical demultiplexers. In this paper, we describe a translucent OBS core node able to perform re-aggregation. This architecture opens the possibility for AQ grooming and is thus expected to reduce the contention rate. The aggregation improvement

Digital Object Identifier: 10.4108/ICST.BROADNETS2009.7869
http://dx.doi.org/10.4108/ICST.BROADNETS2009.7869

compensates for the translucent architecture impairment in terms of delay and network cost. The benefits of partitioning the network into several clusters has been disclosed in [13]: Clusters communicate throughout an intermediate node, called master node. Master nodes can aggregate several bursts toward the nodes of the same cluster. However, intermediate disaggregation is not considered, consequently hindering aggregation grooming. In this paper, we will show the benefits resulting from the use of re-aggregation: It entails a large aggregation grooming potential so that the loss probability, the delay and the network cost are all reduced.

In Section II, we analyze the traffic behavior in a core network. We describe the newly proposed LCH$^+$ model and highlight a consequence of the data plane transparency: In a core network, traffic is completely defined at the emission and can be described by a simplified model that confirms the relevance of the Engset model. In Section III, we describe the emission process (i.e., the aggregation and the MAC layer) and identify two parameters that can be used to exploit the Engset model. In Section IV, we propose our translucent OBS core node model and describe the aggregation grooming that allows, according to experiments presented in Section V a reduction of the loss rate, the delay and the network cost. Conclusions are drawn in Section VI.

II. OBS Traffic in the Core

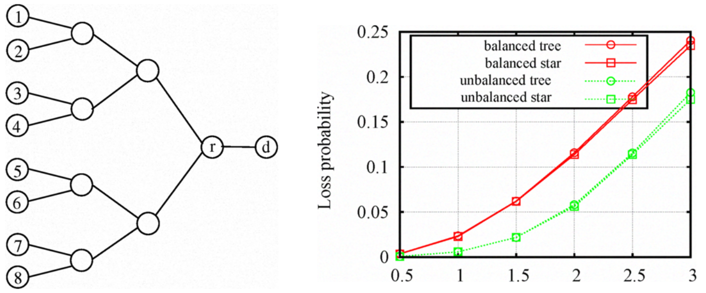
The traffic, as observed inside OBS networks, has several peculiar features, making inappropriate a direct use of most classical results. In this section, we give a survey of these features.

A. *Traffic Properties*

1) OBS Transparency Property: In a buffered network, mixing different flows modifies their characteristics. However, this is not the case in an OBS core network due to the data plane transparency. Indeed, *(i)* the gap between two successive bursts is not affected by a node traversal and sequences of bursts remain unchanged when going through a node – except for some bursts that are discarded due to unresolved collisions ; *(ii)* the travel time between two nodes only depends on the light propagation speed in an optical fiber ($200,000$ km.s^{-1}) and the total traveled distance, whatever the number of traversed switches.

Consequently, the traffic profile on a given link $r \longrightarrow d$ is completely defined at its emission and can be described by a derived star topology where each source is directly connected to r at the same distance as in the original topology.

For a given topology, let us extract the sub-graph that contains only the links carrying traffic toward a given link $r \longrightarrow d$. The inter-arrival between two bursts in $r \longrightarrow d$ depends on their emission date and their respective traveled distance (since they do not necessarily use the same path). Except for dropped bursts, if the same burst emission process is used in the star topology, as the travel distances are identical, the burst arrival dates are the same in both topologies. The reverse is also true. Again, except for dropped bursts, the traffic that is observed in the star topology is equivalent to the traffic observed in the original (mesh) network.



(a) A 3-stage binary tree (b) Loss Probability on Tree and Star Topologies

Fig. 1. Evaluation of the Star Approximation

However, the star topology only provides a lower bound of the loss probability as bursts are not dropped in the same order in both topologies. We have compared the loss probability obtained on the tree topology depicted on Figure 1(a) and its derived star topology with three wavelengths at 10Gbps per link. We consider two traffic scenarios. In the balanced traffic scenario, the overall load is uniformly equally distributed among all sources whereas in the unbalanced one, the load is distributed so that $a_i = 2 \times a_{i-1}$, where a_i is the traffic amount in source i. The initial overall load of 8 Gbps is modified using a multiplicative coefficient (varying between 0.5 and 3) in the different traffic instances (horizontal axis). The loss probability shown on Figure 1(b) reports the accuracy of the star approximation in both balanced and unbalanced scenarios. A very small gap appears under very high load (but then the loss probability is unacceptable). Except for those particular scenarios of limited interest, the star topology can be used as a faithful approximation, and we will go on with star topologies in the sequel.

2) Loss Independent Arrival Property: The burst arrival in a core node is independent of the state of the output ports and the fate of the previous bursts. In addition, bursts submitted by a given input port cannot overlap. As a consequence, the traffic arrival in an OBS core switch can be described by the diagram on Figure 2: The input port either submits a burst of average duration $1/\mu$ or waits τ time units before submitting the next burst. Due to statistical multiplexing, several connections can be superimposed and an input port can submit bursts from a large number of connections. Hence, the burst submission process of an input port is a mix of several arrival processes. In the context of our study, we do not attempt to give a precise account of the arrival process and we simply assume that the silent periods are exponentially distributed, in order to simplify the comparison of various scenarios. The analysis under more general processes is left for a future study.

We denote by $\lambda = 1/\tau$ the average silent period of the input port. The burst submission remains the same, whether the burst is dropped or served. In the example, burst B_1 is dropped, but this event has no impact on the next arrival. In the sequel, an input port is referred to as *active* for the

Digital Object Identifier: 10.4108/ICST.BROADNETS2009.7869
http://dx.doi.org/10.4108/ICST.BROADNETS2009.7869

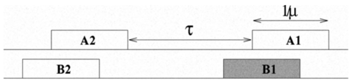


Fig. 2. Core Network Traffic Arrival

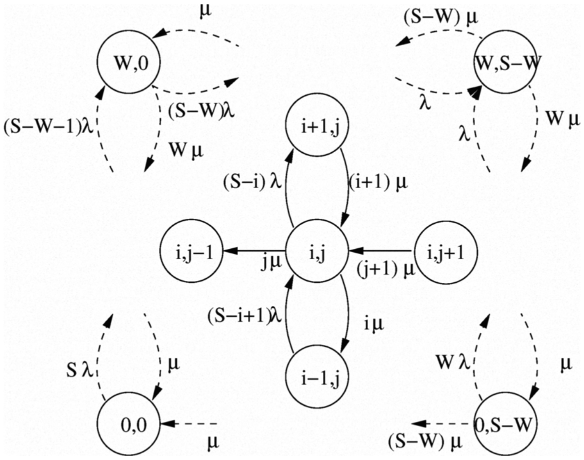


Fig. 3. LCH$^+$ Model

whole expected duration of the burst, whether the burst is successfully transmitted or not.

When we observe a group of N input ports, a direct consequence of the above remark is that the number of active ports is distributed according to a binomial law:

$$P(n) = \binom{N}{n} \alpha^n (1-\alpha)^{N-n}, \qquad (1)$$

where α stands for the load offered by a given source:

$$\alpha = \frac{1}{\mu(1/\mu + \tau)} = \frac{1}{1 + \tau\mu}.$$

B. OBS Core Network Traffic Model (LCH$^+$)

Traditional models assume that requests (packets, requests for connection, etc.) arrive according to a Poisson process where Erlang model applies. In an OBS core network, however, any node multiplexes a finite (usually small) number of incoming links and Erlang model must be replaced by models that take profit of the finite number of sources such as the streamline effect [7] or the Engset model (see, e.g., [14], [15]). Those models do not reflect the independent arrival property. The model proposed in [16] can be easily adapted to reflect the independent arrival property, but describes Poisson process. As far as we know, only the "Lost Call Held" model (LCH) [15], [17] combines the finite number of sources and the "independent arrival property", but it unfortunately describes a system where segmentation is performed [18]. We describe below a variation (LCH$^+$) that discards the burst segmentation. We assume exponentially distributed burst sizes and define $\alpha_i = 1/(1 + \tau_i \mu_i)$ and $\lambda_i = 1/\tau_i$ accordingly to Section II-A2.

Because of the independent arrival property, the state of an input port is independent of the state of the system. In addition, the transparency assumption implies that if a burst is served, the input port and the output port remains both active as long as the input port keeps on receiving data (i.e., for the whole duration of the burst). Those two properties imply that an input port is active for the whole expected duration of each burst, whether the burst is served or dropped.

Figure 3 describes the model of an outgoing link with W output ports (servers) and S identical input ports (sources). In state (i, j), i servers are busy (and associated with i input ports currently receiving data) and j bursts are being dropped (i.e., j input ports are currently receiving data to be dropped). Consequently, $i + j$ sources are active and $S - (i + j)$ sources (qualified as *idle*) can submit a new burst (the birth rate is $(S - i - j)\lambda$). If a wavelength is available ($i < W$), the next burst is served and the system switches to state $(i + 1, j)$ (the input port is connected to a server and remains active for the duration of the burst). Otherwise ($i = W$), the burst is dropped and the next state is $(W, j + 1)$. Here again, the input port remains active as long as it keeps on receiving bursts, although they are dropped (in opposition to an Engset system where an input port would stay *idle*).

When an input port stops receiving data, it becomes *idle* ($i + j$ is decreased by one). If its burst has been dropped (rate $j\mu$), a "dropped input port" becomes *idle* and the next state is $(i, j - 1)$. Otherwise (rate $i\mu$), a "served input port" becomes *idle* and a server is released, leading to state $(i - 1, j)$. Such a behavior differs from the LCH model where a server that becomes available can serve the remaining of a burst being dropped (the system switches to state $(i, j - 1)$ if $j > 0$ and to state $i - 1, 0$ otherwise).

The solution of the LCH$^+$ model provides the probability $P_{i,j}$ for the system to be in state (i, j). A client is rejected if it is submitted while the system is in any state (W, j). The loss probability is the ratio of the rejection rate over the total submission rate:

$$\mathrm{LCH}^+(\alpha, N, W) = \frac{\sum_{i=0}^{N-W-1} (N - (W + i)) P_{W,i}}{\sum_{w=0}^{W} \sum_{i=0}^{N-W-1} (N - (w + i)) P_{w,i}}. \qquad (2)$$

C. Impact of a Finite Number of Incoming Flows

Consider an optical link ℓ with two wavelengths at 10Gbps. Let an overall incoming load of 4 Gbps be equally distributed among S input ports requesting ℓ and following the behavior described in Section II. The loss probability predictions reported on Figure 4 validate the relevance of the LCH$^+$ and Engset models whereas the streamline formula [7] quickly converges toward the Erlang-B loss formula that overestimates the loss.

This simple experiment illustrates that, at equal load, it is beneficial to reduce the number of sources. In our study, a source is an input port. Assuming full wavelength conversion, permutation of the wavelengths in the wavelength assignment leads to an equivalent configuration and, if we systematically select the available wavelength with the lowest index, the

Digital Object Identifier: 10.4108/ICST.BROADNETS2009.7869
http://dx.doi.org/10.4108/ICST.BROADNETS2009.7869

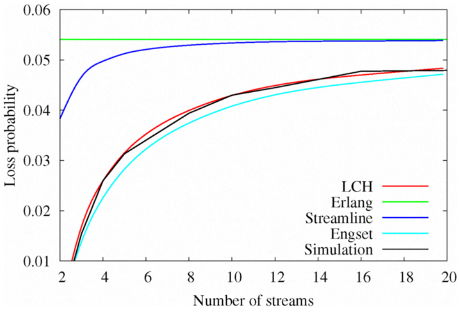


Fig. 4. Influence of the Number of Sources

number of sources seen by an optical link ℓ equates the maximum number of overlapping bursts requesting ℓ (its overlapping occurrence value is denoted by Ω_ℓ). Thanks to the data plane transparency, Ω_ℓ can be controlled at the emission.

III. Burst Emission

In this section, we describe the aggregation and the MAC processes and investigate how the emission process impacts the overlapping occurrence in a core network.

A. Aggregation

Aggregation takes place in aggregation queues (AQs) managed by ingress nodes. Each AQ is assigned a particular combination of traffic characteristics – usually an optical egress node and possibly a class of service – and stores the incoming IP packets that match with those characteristics. Once the aggregation is completed, the packets in the AQ are grouped together into a burst and sent to the MAC layer so that they are all signaled by a single header.

The burst is triggered either once the AQ reaches a given size or after a timer has expired. The former solution supplies bursts of equal sizes but the aggregation duration increases for low loaded AQs. The latter solution bounds the aggregation duration, but low loaded AQs generate small bursts and increase the signaling overhead. A combination of both criteria is commonly accepted as a reasonable compromise [2]. Several variations have been proposed to improve the aggregation process (e.g., [3]) but all basically conform with the described mechanism.

Delay consideration apart, several contributions argued in favor of bursts with a fixed size (e.g., [19]–[21]). We assume here that all bursts have the same size Np, where p is the size of the packets arriving according to a Poisson process and N is the number of packets. The service rate is denoted by $\mu = C/(Np)$ where C is the transport capacity of the wavelengths. The mean aggregation duration amounts to N/λ and the bursts are submitted to the MAC layer at a rate $\Lambda = \lambda/N$ following an Erlang-N distribution.

$$f(x, N, \lambda) = 1 - \sum_{n=0}^{N-1} e^{-\lambda x}(\lambda x)^n/n!. \quad (3)$$

B. Medium Access

The MAC layer is responsible of the wavelength assignment and the scheduling of newly aggregated bursts. Denoting t_{READY} the date the burst is aggregated and OT its offset time, the medium access controller looks for a wavelength available at time $t_{\text{READY}} + OT$ for the whole duration of the burst.

At that point, however, the burst emission can be postponed to meet the emission objective. Especially, the MAC layer can serialize the emission of the bursts to limit the overlapping degree to a maximum value $\Omega^{\max}$. In that case, a burst is delayed if, at its arrival in the MAC, $\Omega^{\max}$ or more bursts are being emitted. Such events increase both the medium access delay and the required buffering capacity.

C. Observation of the Emission Process

1) Impact of the Aggregation Process: The bursts are aggregated according to an Erlang-N distribution whose CDF (Cumulative Distribution Function) is plotted in Figure 5(a) with packets arrival rate $\Lambda = 1$ and various bursts size N. Two bursts overlap if their inter-arrival is shorter than their duration ($1/\mu = Np/C$), i.e., with probability $f(1/\mu, N, \lambda)$. Figure 5(b) plots such a probability regarding N and the load of the AQ. It reports that the overlapping probability decreases with larger N as confirmed in Figure 5(c) that plots the overlapping degree (Ω_ℓ) measured for similar experiments. Figure 5(c) also reveals that, for a given load, Ω_ℓ is reduced if less AQs are involved. For example, with $N = 10$, up to three bursts will overlap if a load of 1.6 Erlang is handled by a single AQ, versus 4, 4 and 16 with respectively 2, 4 and 8 equally loaded AQs.

Let us now evaluate the impact of the aggregation on the loss probability and the delay. The simulation involves two ingress nodes connected to a core node v. Each ingress node generates a load of 1.6 Erlang competing for four wavelengths in v. Incoming packets are uniformly spread among 1, 2, 4, 8 or 16 AQs that generate bursts of size N packets. We assume that the MAC schedules the burst emissions as soon as possible. The loss probability plotted on Figure 5(e) confirms the positive impact of the reduction of the overlapping degree on the loss probability. Concentrating AQs systematically reduces the loss and reduces the aggregation delay (as reflected by the end-to-end delay reported on Figure 5(f)). Increasing the size of the bursts helps reducing the loss rate, but the improvement decreases when increasing N whereas the end-to-end delay continuously increases with N. Note that the increase of the aggregation delay induced by longer bursts can be compensated by grooming AQs.

2) Impact of the MAC policy: Aggregation process apart, the overlapping degree can be controlled by the MAC. In particular, the MAC can be configured to restrict the number of wavelengths simultaneously used by the bursts of a given AQ. This method impacts the time spent in the MAC buffers and consequently the buffering capacity requirements. Figure 5(d) plots the average medium access delay regarding the load of the AQ and the overlapping restriction $\Omega^{\max}$ for $N = 10$. It suggests the relevance of using aggregation to shape the traffic.

Digital Object Identifier: 10.4108/ICST.BROADNETS2009.7869
http://dx.doi.org/10.4108/ICST.BROADNETS2009.7869

For example, with $\Omega^{\max} = 2$, bursts generated by a single AQ load at 1 Erlang will spend about 0.1ms in the MAC whereas they will spend 0.7ms if the load is spread among two equally loaded AQs restricted to one wavelength each.

IV. Intermediate Re-aggregation and Aggregation Grooming

To generate traffic that conforms with the favorable profile identified in Section II, we propose to feature re-aggregation in OBS core nodes. This architecture offers access to electrical buffers and opens the possibility to aggregation queue grooming that should improve the aggregation process and decrease the overlapping degree in core networks.

A. Re-aggregation

Intermediate re-aggregation is close to grooming in OCS, carried out by MSPPs in the electrical domain: A connection is not necessarily served all optically from its origin to its destination, but can be subject to electrical processing at some intermediate nodes.

Figure 6 presents the slight modification of the OBS node architecture that can be found, e.g., in [22]. The node operates as an edge node – receiving incoming data from ADD ports to be sent into the network – and as a core node – receiving optical bursts from other nodes of the network. The new incoming traffic is sent to the aggregation module. The bursts arriving from optical ports either cut-through the SOA (Semiconductor Optical Amplifier) array toward the next OBS node or are directed to the disaggregation module. In the latter case, we propose to make it possible to re-inject part of the disaggregated burst into the aggregation module.

Encapsulation of each packet can be considered to keep the OBS layer independent of the packet format. By attaching a label to each packet to specify its final destination, the disaggregation module can easily differentiate the packets to be sent to the DROP port (those whom final destination is reached) from those to be resubmitted to the aggregation module. The labels can however be coded on few bits and the involved overhead should not impact the resource utilization efficiency.

The re-aggregation offers access to electrical buffering and can be seen as a reactive solution to the loss issue in OBS. In the next section, we will describe how it can be used to proactively help reducing the loss probability. Note that additional benefits can be expected regarding, e.g., the reduction of the offset based priority (see [1] for more details) and the retransmission process (as retransmission is performed by the last aggregation node, the notification delay and the amount of resources wasted by a dropped burst are reduced).

We next discuss two possible drawbacks. Firstly, the re-aggregation could increase the cost of the network. One can expect a fixed cost reflecting the additional connections inside the node and a variable cost induced by additional electrical buffering capacity requirements.

Secondly, re-aggregation implies additional aggregation, disaggregation and medium access procedures and may increase the related delays (denoted δ_{AGG}, δ_{DEAGG} and δ_{MAC}). The buffer requirements and the delay increases are experimentally evaluated in Section V.

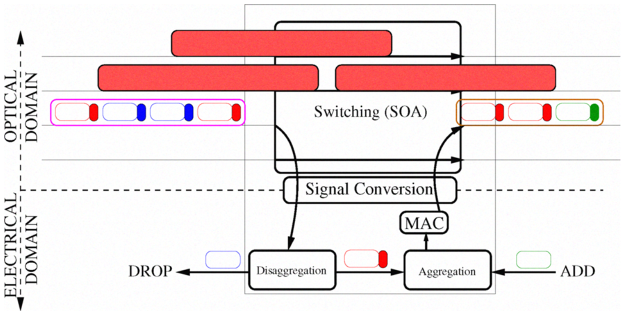


Fig. 6. Re-aggregation Capable Node Architecture

B. AQ grooming

At a given ingress node, packets with different destinations can be mixed in the same burst if re-aggregation is performed somewhere during its journey. In other words, several aggregation queues can be merged. It can can be exploited in two ways: *(i)* The resulting AQ is more loaded, speeding up the aggregation process ; *(ii)* The reduction of the number of AQs reduces the overlapping occurrences (see Section III-C) which is favorable from the core network point of view (see Section II) and simplifies the MAC procedures.

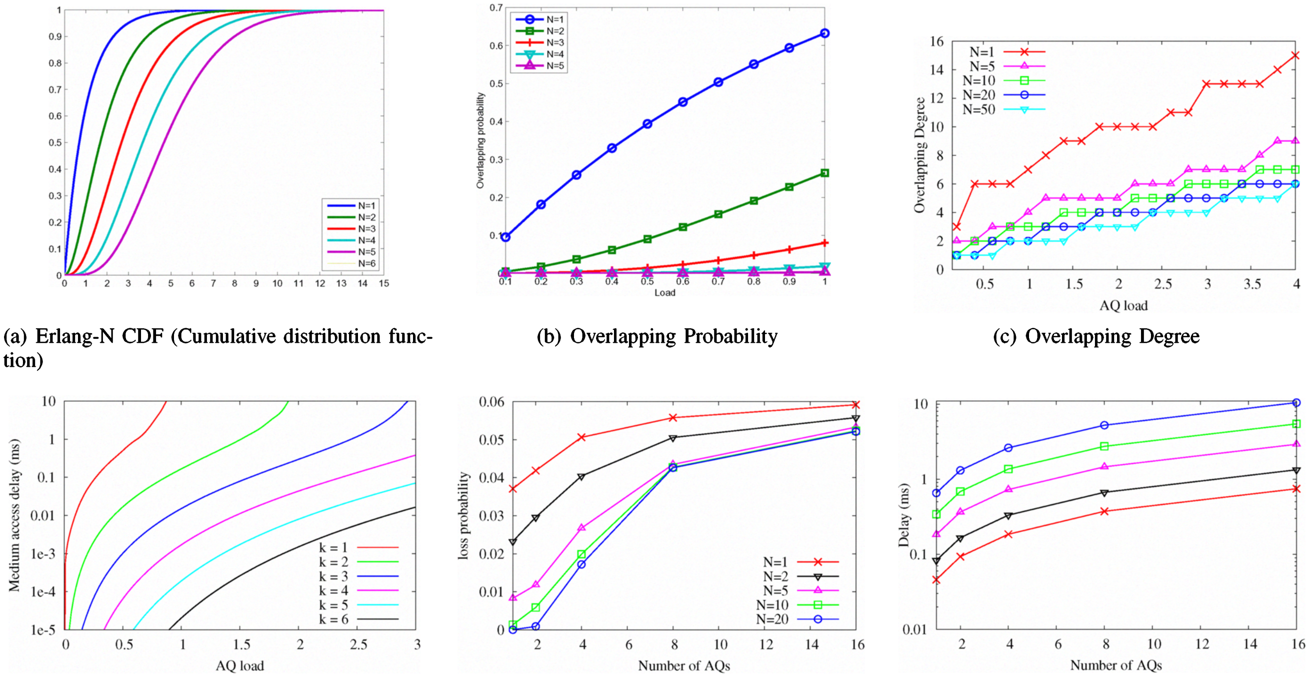
On Figure 7(a) and Figure 7(b), we assume traffic from nodes v_1 and v_2 to nodes v_3 and v_4. $C_{i,j}$ denotes the connection from v_i to v_j. With end-to-end aggregation (Figure 7(a)), each ingress node manages one AQ per destination and the bursts compete in switch S_1 to reach S_2. If re-aggregation is featured in S_2 (Figure 7(b)), ingress node v_i only maintains one AQ that mixes packets of $C_{i,3}$ and $C_{i,4}$ in bursts sent to S_2. The surviving bursts are disassembled in S_2, that maintains one AQ per destination, regardless of the origin.

Consider the example of Figure 7 with $N = 10$ and connections load of 0.5. The maximum number of overlapping bursts arriving in S_1 (denoted by Ω_{S_1}) is eight (two for each connection according to Figure 5(c)) with end-to-end aggregation. With re-aggregation featured in S_2, aggregation grooming allows the reduction of Ω_{S_1} to 6 (three per flow from v_i to S_2). Simulations run under the assumption of four wavelengths per link, confirm the correlation between the overlapping degree and the loss rate (Figure 5(e)): Increasing the burst length and reducing the number of AQs contribute to reduce the loss rate. Figure 5(f) illustrates the benefits of aggregation grooming regarding the end-to-end delay (from 6.10^{-3} ms to 8.10^{-4} ms): Merging AQs increases the load of the resulting AQs and reduces the aggregation delay. Note that the reduction of the overlapping degree should reduce the probability to delay a burst in the MAC and consequently reduces the medium access delay.

C. Grooming Strategy

We conducted the experiments with the use of a greedy AQ grooming strategy. For a given routing configuration and given

Digital Object Identifier: 10.4108/ICST.BROADNETS2009.7869
http://dx.doi.org/10.4108/ICST.BROADNETS2009.7869



(a) Erlang-N CDF (Cumulative distribution function)

(b) Overlapping Probability

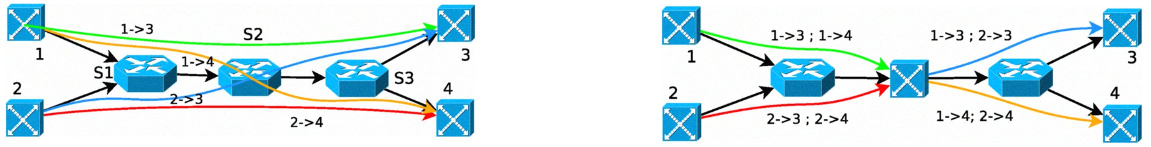
(c) Overlapping Degree

(d) Medium Access Delay with Overlap Restriction

(e) Loss

(f) Delay

Fig. 5. Burst Emission Performances



(a) End-to-End Aggregation

(b) Intermediate Re-aggregation

Fig. 7. Re-aggregation Example

the set $\mathcal{R}$ of re-aggregation capable nodes, we groom as many AQs as possible so that the routes are not modified. Thus, if $v \in \mathcal{R}$, AQs $A_1^{v'}$ and $A_2^{v'}$ managed in node v' are groomed if their bursts reach v_r via a common sub-path. Consequently, every packet that uses the same path from a node v' to v are handled by a single AQ.

In v, the content of the disaggregated bursts is sorted among AQs depending on the next disaggregation node.

V. Experimental Results

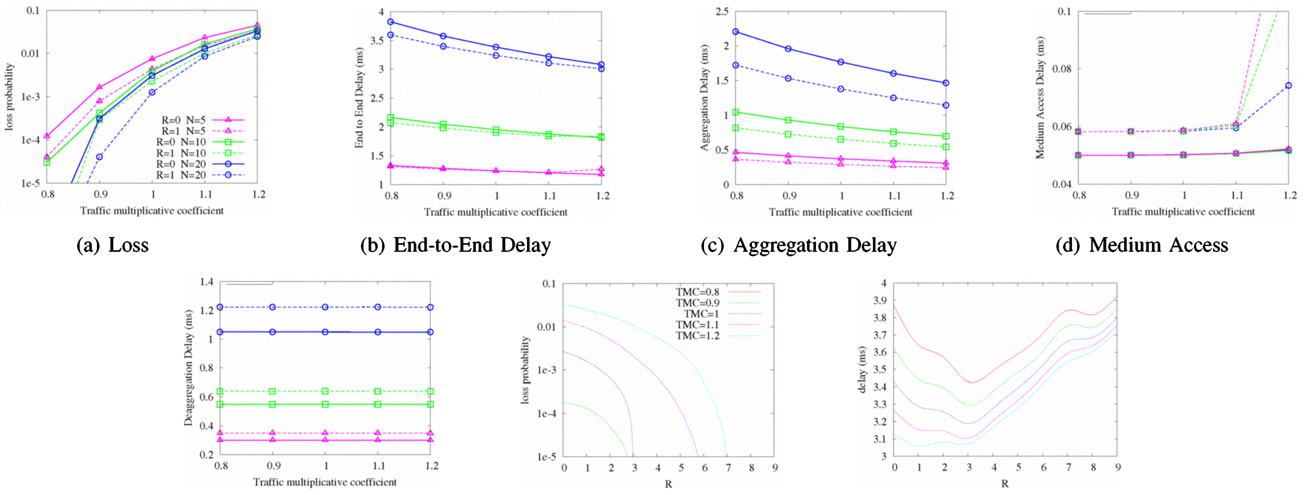
We use the EONET topology and the traffic matrix found in [23] with $C = 10$ Gbps. The overall load is tuned with a multiplicative coefficient. Full wavelength conversion is featured in every node using equipment described in [24], [25] so that the conversion delay can be neglected. We assume that Opto-electrical conversion is performed at the same rate as the transport capacity 0.1 ms.Mb^{-1}. The header processing time is set to $50\mu s$ ([22]) whereas the guard time can be neglected above hundreds of nano-seconds (the switch reconfiguration time is evaluated to few nanoseconds in [26]). Incoming packets have equal sizes (1 Mb) and arrive according to a Poisson distribution.

A. Performance on Plain Network

In this first set of experiments, W is set to 30 and $\mathcal{R}$ is a set of R re-aggregation capable nodes. The routing configuration is obtained with the model proposed in [27] (it aims to minimize the load of the bottleneck). The re-aggregation capable nodes are selected so that the amount of groomed flows (with the grooming strategy of Section IV-C) is maximized.

Figure 8(a) reports the loss probability with $R \in \{0, 1\}$. First note that, as observed in Section IV-B, increasing the burst length reduces the loss probability: With $R = 0$, increasing N from five to 10 and 20 packets successively reduces the loss probability from 16.4×10^{-4} to 4.1×10^{-4} and 3.1×10^{-4} with a traffic multiplicative coefficient of 0.9. The benefits are reduced with higher loads, but remains significant (from 0.023 to 0.017 and 0.013 with a traffic multiplicative coefficient of 1.1). The re-aggregation ($R = 1$) reduces further the loss probability. With $N = 10$, the loss probability is reduced from 25% under high load up to 80% under low load.

Digital Object Identifier: 10.4108/ICST.BROADNETS2009.7869
http://dx.doi.org/10.4108/ICST.BROADNETS2009.7869



(a) Loss

(b) End-to-End Delay

(c) Aggregation Delay

(d) Medium Access

(e) Disaggregation Delay

(f) R versus the Loss (N=20)

(g) R versus the Delay (N=20)

Fig. 8. Experimental Results on plain EONET

The delay (Figure 8(b)) increases with the use of longer bursts, but is reduced if the load increases or if re-aggregation is used. It fluctuates in the same direction as the aggregation delay (see Figure 8(c)) because the aggregation process plays a major role in the end-to-end delay. If the load increases, the aggregation delay is reduced and impacts the end-to-end delay. Similarly, the re-aggregation allows AQ grooming and also contributes to reduce the aggregation delay and compensates for the increase of the disaggregation (Figure 8(e)) and medium access delay (Figure 8(d)) that are performed more often. Note that the contribution of the MAC delay becomes significant under high load (traffic multiplicative coefficient $\geq$ 1.2). It is however less impacted with longer bursts, because the reduction of the overlapping degree increases the wavelength availability.

The use of re-aggregation can also impair the delay if R increases above a given limit (Figure 9(a)). The loss probability is systematically reduced by additional re-aggregation capable node thanks to the effect of the AQ grooming on the traffic profile and the better access to electrical buffering.

B. Performance on a Clustered Network

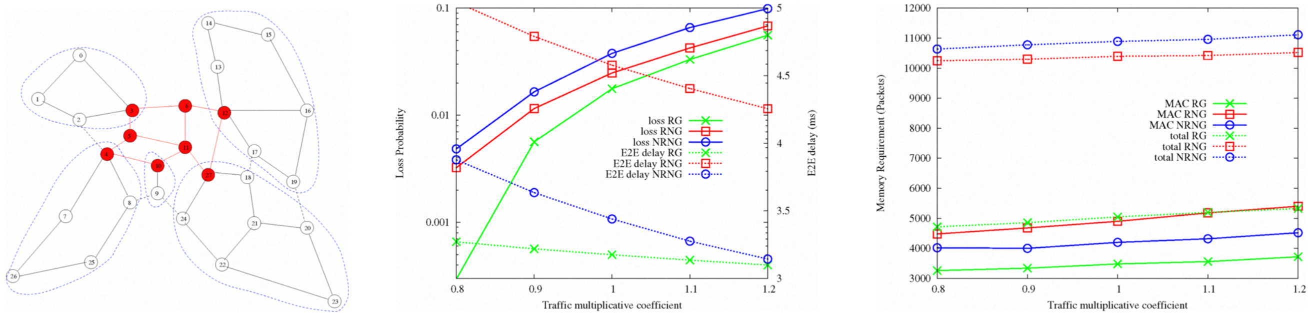
In [13], the authors show the benefits of partitioning the network into several clusters. Clusters communicate via a master network in which each node acts as a gateway for a given cluster. The master nodes can aggregate several bursts together in the upstream direction but do not disassemble bursts so that aggregation grooming cannot be used.

Although full re-aggregation in the master nodes increases the complexity of their aggregation process, it offers a high aggregation grooming potential. Figure 9(a) presents a clustered EONET topology. Eight clusters are defined, linked by their master nodes that perform re-aggregation of the traffic from / to its cluster (i.e., at the entrance and exit of a cluster). Consequently, a non-master node can mix any data to be sent towards its master node or outside its cluster. Master nodes manage one AQ for each node of its cluster (downstream) and one AQ for each master node (upstream). We compare the electrical buffering requirement (Figure 9(c)), the loss probability and the delay (Figure 9(b)) of three configurations: The classical OBS-JET (no re-aggregation and no aggregation grooming, NRNG), the clustered architecture with re-aggregation but without aggregation grooming (RNG) and the clustered architecture with both re-aggregation and aggregation grooming (RG).

The simple use of the re-aggregation (RNG vs. NRNG) reduces the loss probability thanks to the electrical buffering in the gateways. The counter-part of re-aggregation is the significant increase of the end-to-end delay (around 30%) and the need for about 20% more electrical buffer resources in the MAC. The aggregation grooming (RG) allows overcoming those drawbacks and achieves the best of NRNG and RNG: The high grooming potential resulting from full re-aggregation overrides the delay impairment due to electrical buffering and globally reduces the end-to-end delay beyond OBS-JET. The reduction is more visible under low load since the burst aggregation delay is severely increased if aggregation grooming is not performed. The use of aggregation grooming also significantly reduces the loss rate as compared with simple re-aggregation. This illustrates the benefits earned by reducing the overlapping degree. Finally, the aggregation grooming reduces the memory requirement by half, firstly because the number of AQs is reduced, but also because, the bursts are less likely overlaping and spend less time in MAC buffers.

VI. Conclusions

The buffer-less nature of OBS implies a particular behavior in a core network that is accurately described with the LCH$^+$ model. The data plane transparency has then been exploited to propose a simplified loss approximation technique that discloses the importance of the emission process: It completely defines the traffic profile in the core. The relevance of the



(a) Clustered topology

(b) Loss rate and delay

(c) Memory requirement

Fig. 9. Experimental results on a clustered EONET

Engset model in the core suggests to reduce the overlapping degree at the emission. We identified two impacting parameters of the aggregation process: The number of aggregation queues and the burst length. To sidestep the topology constraint that governs the aggregation queue definition, we proposed a translucent architecture able to perform intermediate re-aggregation. This architecture offers access to electrical buffering and opens the possibility to aggregation grooming to reduce the aggregation delay (so that the bursts can be lengthened transparently) and the overlapping degree. In addition to the significant reduction of the loss rate, experiments run on the EONET network revealed that the improvement of the aggregation earned by aggregation grooming can hide the negative impact of electrical buffering on the delay. We envisioned the re-aggregation in cluster interconnection nodes and showed that the high aggregation grooming potential in such a case allows to reduce the loss rate, the delay and the network cost.

REFERENCES

[1] M. Yoo, M. Jeong, and C. Qiao, "A new optical burst switching (OBS) protocol for supporting quality of service," in *SPIE*, vol. 3531, November 1998, pp. 396–405.

[2] X. Cao, J. Li, Y. Chen, and C. Qiao, "Assembling TCP/IP packets in optical burst switched networks," in *IEEE Global Telecommunications Conference - GLOBECOM*, vol. 3, November 2002, pp. 2808–2812.

[3] T. Hashigushi, X. Wang, H. Morikawa, and R. Aoyama, "Burst assembly mechanism with delay reduction for OBS networks," in *Proceedings of COIN/COFT*, 2003, pp. 664–66.

[4] V. Vokkarane, Q. Zhang, J. Jue, and B. Chen, "Generalized burst assembly and scheduling techniques for QoS support in optical burst-switched networks," in *IEEE Global Telecommunications Conference - GLOBECOM*, vol. 3, 2002, pp. 2747 – 2751.

[5] A. Zalesky, H. Vu, Z. Rosberg, E. Wong, and M. Zukerman, "OBS contention resolution performance," *Performance Evaluation*, vol. 64, no. 4, pp. 357–373, 2007.

[6] T. Coutelen, B.Jaumard, A. Metnani, and H. Elbiaze, "Using network load and traffic load balancing for an efficient deflection routing scheme," in *IASTED Proceedings of Optical Communications Systems and Networks (OCSN)*, Banff, Alberta, Canada, July 2005.

[7] M. Phùng, K. Chua, G. Mohan, M. Motani, and T. Wong, "The streamline effect in OBS networks and its application in load balancing," in *2nd International Conference Broadband Networks*, October 2005, pp. 304 – 311.

[8] C. Gauger, "Dimensioning of FDL buffers fr optical burst switching nodes," in *Conference on Next Generation Optical Network Design and Modelling - ONDM*, 2003, pp. 117–132.

[9] G. Gurel and E. Karasan, "Using multiple per egress burstifiers for enhanced TCP performance in OBS networks," *Photonic Network Communications*, vol. 17, pp. 105–117, April 2009.

[10] *Synchronous Optical Burst Switching*, 2004.

[11] F. Farahmand, Q. Zhang, and J. Jue, "Dynamic traffic grooming in optical burst-switched networks," *Journal of Lightwave Technology*, vol. 23, no. 10, pp. 3167–3177, 2005.

[12] Y. Fan and B. Wang, "Exploring node light-splitting capability for burst grooming in optical burst switched networks," *Photonic Network Communications*, vol. 14, no. 2, pp. 209–222, 2007.

[13] J. Angelopoulos, K. Kanonakis, G. Koukouvakis, H. Leligou, C. Matrakidis, T. Orphanoudakis, and A. Stavdas, "An optical network architecture with distributed switching inside node clusters features improved loss, efficiency, and cost," *Journal of Lightwave Technology*, vol. 25, pp. 1138 – 1146, May 2007.

[14] G. Fiche and G. Hébuterne, *Communicating Systems & Networks: Traffic & Performance*. Kogan Page Science, 2004.

[15] R. Syski, *Introduction to Congestion Theory in Telephone Systems*. Elsevier, 1986.

[16] Q. Zhang, V. Vokkarane, J. Jue, and C. Chen, "Absolute QoS differentiation in optical burst-switched networks," *IEEE Journal on Selected Areas in Communications*, vol. 22, pp. 1781 – 1795, 2004.

[17] A. Viterbi, *CDMA Principles of spread spectrum communication Addison-Wesley wireless communications series*. Addison-Wesley Pub, 1995.

[18] V. Vokkarane, J. Jue, and S. Sitaraman, "Burst segmentation: An approach for reducing packet loss in optical burst switched networks," in *IEEE Global Telecommunications Conference - GLOBECOM*, vol. 5, 2002, pp. 2673–2677.

[19] X. Yu, Y. Chen, and C. Qiao, "A study of traffic statistics of assembled burst traffic in optical burst switched networks," in *Proceedings of Opticomn*, vol. 4874, 2002, pp. 149–159.

[20] J. Li, C. Qiao, J. Xu, and D. Xu, "Maximizing throughput for optical burst switching networks," *IEEE Annual Joint Conference of the IEEE Computer and Communications Societies - INFOCOM*, vol. 3, pp. 1163–1176, 2004.

[21] ——, "Maximizing throughput for optical burst switching networks," *IEEE/ACM Transactions on Networking*, vol. 15, pp. 1163–1176, October 2007.

[22] Y. Sun, T. Hashiguchi, V. Q. Minh, X. Wang, H. Morikawa, and T. Aoyama, "Design and implementation of an optical burst-switched network testbed," *IEEE Communications Magazine*, vol. 43, pp. S48–S55, November 2005.

[23] A. Betker, C. Gerlach, R. Hulsermann, M. Jager, M. Barry, S. Bodamer, J. Spath, C. Gauger, and M. Kohn, "Reference transport network scenarios," MultiTeraNet Project, Tech. Rep., 2004.

[24] *Error-free 320 Gb/s SOA-based Wavelength Conversion using Optical Filtering*, 2006.

[25] Y. Liu, E. Tangdiongga, Z. Li, H. de Waardt, A. Koonen, G. Khoe, X. Shu, and H. Dorren, "Error-free 320-Gb/s all-optical wavelength conversion using a single semiconductor optical amplifier," vol. 25, pp. 103–108, January 2007.

[26] T. El-Bawab and J. Shin, "Optical Packet Switching in Core Network : Between Vision and Reality," *IEEE Communications Magazine*, vol. 40, no. 9, pp. 60–65, 2002.

[27] T. Coutelen, "Accès et routage optique en mode de commutation de rafales," Master's thesis, Université de Montréal, Montréal, Canada, 2005.

Digital Object Identifier: 10.4108/ICST.BROADNETS2009.7869
http://dx.doi.org/10.4108/ICST.BROADNETS2009.7869