Ingress Traffic Conditioning in Slotted Optical Packet Switched Networks

Vijay Sivaraman, David Moreland and Diethelm Ostry
ICT Centre, CSIRO
Epping, NSW 1710, Australia
Email: {Vijay.Sivaraman, David.Moreland, Diet.Ostry}@csiro.au

Abstract—**Optical Packet Switched (OPS) networks are known to suffer from high packet losses due to contentions. Techniques such as fiber delay line (FDL) buffering and wavelength conversion help to relieve contentions to an extent, but are expensive and are therefore used sparingly. This paper explores how electronic edge nodes can *condition* the traffic entering the OPS network with the objective of utilising the few available FDLs (or wavelength converters) more effectively to reduce contention losses further within the all-optical core.**

The benefits of ingress traffic conditioning in electronic networks are well known, but we believe its advantages for OPS networks with sparse contention resolution resources have not been fully evaluated. This paper considers a time-slotted OPS network comprising core nodes with limited FDL buffering. The edge-nodes are allowed to condition the ingress traffic by controlling the release of packets into the optical network in accordance with a particular scheme. In this context, our contributions are two-fold. First, we show analytically that for an optical switch with given FDL buffering capacity, contention losses on its output link can be minimised by *evenly* spacing packets on that link. This motivates edge traffic conditioning as an effective means of reducing contention losses in the OPS core. However, ingress traffic conditioning incurs a penalty, namely an increased queueing delay at the edge node. As our second contribution, we quantify via simulation the reduction in packet losses and the accompanying increase in end-to-end delay as the traffic conditioning varies. Our simulated OPS network topology is based on an actual trans-continental Australian network called CeNTIE (a research network in the same vein as Internet2 and Canarie), modeled with realistic traffic flows carrying long-range dependent traffic. We believe our proposal can assist a network operator to trade-off delay for loss when choosing an operating point for the OPS network.

I. Introduction

The maturing of Wavelength Division Multiplexing (WDM) technology in recent years has made it possible to harness the enormous bandwidth potential of an optical fibre cost-effectively. As systems supporting hundreds of wavelengths per fibre with transmission rates of 10-40 Gbps per wavelength become available, electronic switching is increasingly challenged in scaling to match these transport capacities. All-optical switching [1] is widely recognised as the key to meeting these challenges. All-optical switching seeks to eliminate the electronic bottleneck, while dramatically lowering system cost by minimising opto-electronic conversion.

Much initial work on all-optical switching focused on wavelength routed networks supporting end-to-end lightpaths with the full bandwidth of a wavelength [2], [3]. Such psuedo-statically provisioned networks are however inflexible and inefficient for the transport of data traffic, which is inherently bursty. To meet the diverse demands of data traffic, next-generation networks will need to support statistical multiplexing and fine-grained (sub-wavelength) bandwidth allocation. Several approaches to optical subwavelength switching have been proposed in the literature (e.g. [4], [5]), among which Optical Packet Switching (OPS) [6] is attracting increasing attention. Several experimental test-beds [7], [8], [9], [10], [11] have recently demonstrated the feasibility of OPS.

A fundamental concern in OPS networks is contention, which occurs at a switching node whenever two or more packets try to leave on the same output link, on the same wavelength, at the same time. In electronic store-and-forward switches, contention is resolved relatively easily by buffering packets in RAM. In optical packet switches, however, the only practical optical buffers available today are fibre delay lines (FDLs) [12], which are too expensive to implement the large delays needed for acceptable loss performance. Other contention resolution mechanisms, such as wavelength conversion [13], [14], deflection routing [15] and combinational schemes [16] have emerged in the literature. However, these schemes also have limitations (high cost, packet reordering, etc.), are not considered in this paper, though our work does not preclude their use.

Given that practicable optical packet switches are likely to have only limited FDL buffering capacity, we investigate in this paper how the edge nodes can assist the optical network in using these FDLs most effectively to minimize contention losses. Our proposition is based on the observation that when multiple traffic streams are interleaved at an optical switch, less FDL buffering is required to achieve desired losses if the packets in each stream are spaced apart from each other rather than bunched together. Based on this premise, we evaluate how ingress traffic conditioning, namely the regulation of packet spacing into the optical network by the electronic edge node, helps reduce OPS network packet losses. Traffic conditioning has been studied extensively in the context of electronic networks, mainly with the objective of making the loss-delay performance within the network more predictable. Our framework developed in this paper is specifically adapted to optical networks, wherein the scheduling at a switch is time-constrained (by FDL delaying capacity), as opposed to electronic switches where scheduling is space-constrained (by RAM capacity).

Our contributions in this paper are two-fold. First, we establish by analysis that for an optical switch with given FDL buffering capacity, contention losses on its output link can be minimised by *evenly* spacing packets on that link. This motivates ingress conditioning as a means of reducing contention losses in the OPS network. However, conditioning incurs a cost, namely increased queueing delay at the ingress. Our second contribution, therefore, quantifies the increase in the per-flow end-to-end delay that accompanies the reduction in packet losses as the packet spacing at the ingress is varied. This loss-delay tradeoff is studied via simulations using the topology of an ac-

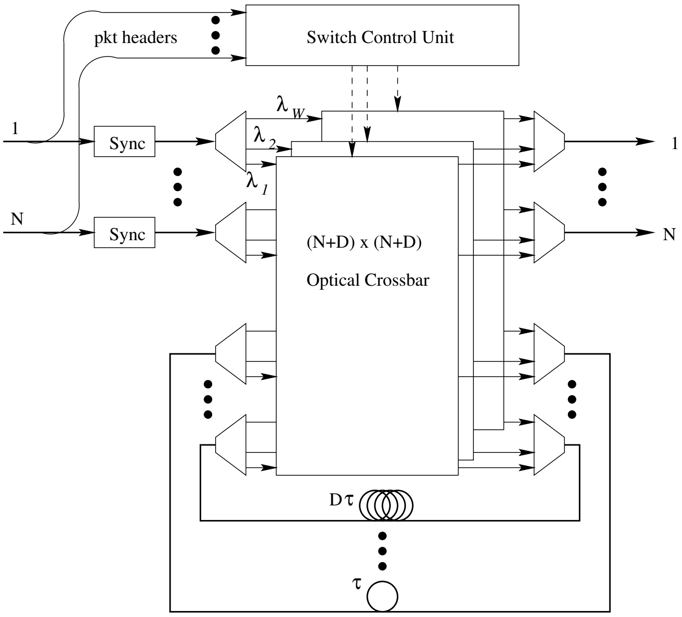

Fig. 1. OPS switch architecture

tual Australian network called CeNTIE, modeled with realistic traffic flows carrying short and long-range dependent traffic.

The rest of this paper is organised as follows: Section II describes the OPS system architecture used in this work. Section III provides the analytical motivation for the ingress traffic conditioning scheme in this OPS context. Section IV demonstrates via simulation the benefits of ingress conditioning in a single-switch setting, while section V quantifies the loss-delay trade-offs in a realistic network scenario. Conclusions and directions for future work are presented in Section VI.

II. System Architecture

There are numerous architectures proposed for OPS networks (see [6], [17] for an overview) – for example, they can be slotted or unslotted, can employ space-switching or broadcast-and-select fabrics, may have feed-forward or feed-back FDLs, etc. This paper investigates traffic conditioning at the electronic *ingress* to an OPS network, and can be applied to any OPS architecture. The system architecture we chose in this paper is described next.

A. Core Node

We use a slotted system [8], [11], with fixed-size optical packets that fit in a slot. Slotted systems typically have lower contentions than unslotted ones [16], and permit synchronous reconfiguration of the switching fabric on a slot-by-slot basis. Figure 1 shows the architecture of the OPS switch. On each of the N input fibres, an optical splitter splits a small amount of power from the incoming packets and sends it to the control unit. The control unit extracts timing information (used for configuring the synchronisation stages) and packet header information (used for determining the packet route and configuring the optical crossbar accordingly). Input signals are synchronised to align packets to slot boundaries, and demultiplexed into the W component wavelengths. We assume that wavelength converters are not employed, and each wavelength traverses its own switching plane. Output port contentions are resolved using a set of D FDL buffers of increasing length that provide delays of $1, 2, \ldots, D$ slots for all wavelengths. This architecture is known as the shared-memory optical packet switch [18], [6].

The buffering architecture presented above permits multiple circulations of packets through the FDLs. However, each recirculation through the crossbar and FDL degrades the optical signal [19]. For this reason we assume here that a packet passes *at most once* through any FDL. Furthermore, if there is contention for a timeslot at an output port of the crossbar, preference is given to the packet entering the crossbar from the longest FDL. In the absence of multiple circulations through the FDLs, the packet egressing the longest FDL is the oldest packet to enter the switch, and giving preference to such a packet ensures that packet ordering is maintained. This strategy schedules an incoming packet into the *earliest* available free slot, and this greedy scheduling allows efficient hardware implementation.

B. Edge Node

Each edge node is assumed to have a singe FIFO queue on each output link. The extension to multiple queues, one per QoS-class, is deferred for future work. The electronic edge node receives packets of varying length from multiple ingress interfaces, and assembles these into fixed-size "optical" packets (with their own optical header) for transmission into the OPS network [20]. Much commendable research has gone into optical packet assembly mechanisms that shape the characteristics of the ingress traffic. The effect of setting the optical packet assembly threshold parameters, such as maximum optical packet fill time and maximum number of bytes per optical packet, on the "shaping" of traffic characteristics has been explored [21], [22], [23], as has the effect of the assembly mechanism on network performance [24], [25]. Our work differs from these studies in that we focus on controlling the release of the already assembled packets into the OPS network with the aim of using the sparse FDL buffering resource more effectively to reduce contention losses in the optical domain.

III. Analysis

In this section we provide the anaytical motivation for ingress traffic conditioning in the OPS context. Consider an arbitrary core optical link $\mathcal{L}$ transporting fixed-size packets in the slotted all-optical network. Denote by $\mathcal{S}$ the OPS switch of which $\mathcal{L}$ is an egress, and by D the maximum delaying capacity (in units of slots) at switch $\mathcal{S}$ (namely, it is assumed that $\mathcal{S}$ supports delays of $1, 2, \ldots, D$ slots). We define the *incremental loss* on link $\mathcal{L}$ as follows:

Definition 1: The *incremental loss* δ on link $\mathcal{L}$ is the probability that *one* additional packet traversing switch $\mathcal{S}$ *cannot* be successfully scheduled onto link $\mathcal{L}$.

Unlike total packet losses in a system, which depend on the arrival process (i.e. possible correlations between successive arrivals), the incremental loss defined above is for a single independent arrival, making it easier to analyse. We believe δ is a good practical indicator of link performance since it provides a good measure of how readily an incremental amount of extra traffic (namely one packet) can be supported by the link. In the following we study how δ varies as the packet spacing is changed.

We compute δ as follows. Let successive slots be consecutively numbered, with 0 denoting the start of the system. Let the counting process $X(n)$, $n > 0$ denote the number of packets transported by the link $\mathcal{L}$ in slots $[0, n-1]$. Note that $\forall n : X(n) \leq n$, since a timeslot carries at most one packet. We assume that $X(n)$ has stationary increments, namely, $\forall m \geq 0 : X(n+m) - X(m)$ has the same distribution as $X(n)$. Now define the probability mass function $f^k(n) = P\{X(n) = k\}$. Further, $\rho = E[X(n)]/n$ denotes the average slot-occupancy ratio (i.e. utilisation) on link $\mathcal{L}$. Then we have:

Theorem 1: The incremental loss δ on link $\mathcal{L}$ connected to switch $\mathcal{S}$ with FDL capacity D is given by

$$\delta = f^{D+1}(D+1) \tag{1}$$

Proof: By definition, δ is the probability that a new packet p arriving at switch $\mathcal{S}$ cannot be scheduled onto link $\mathcal{L}$. Say p arrives at random timeslot j. We claim that p is lost iff link $\mathcal{L}$ is busy in all slots $j, \ldots, j+D$. Note first that p has to be scheduled in one of the slots $j, \ldots, j+D$. This is because p arrives no earlier than slot j, and cannot be delayed by the delay lines to any slot further than $j + D$. For the if part, it is easily seen that if slots $j, \ldots, j+D$ are already occupied by other packets, p cannot be scheduled and is lost. Conversely, if any of the slots in $j, \ldots, j+D$ is idle (does not carry a packet), p can be scheduled there by passing it through the appropriate delay line. Thus p is lost iff all slots $[j, j+D]$ are busy, i.e., iff $X(j+D+1) - X(j) = D+1$. Since $X(n)$ has stationary increments, $X(j+D+1) - X(j)$ has the same distribution as $X(D+1)$. The loss probability for p thus equals $P\{X(D+1) = D+1\} = f^{D+1}(D+1)$. This completes the proof. □

Theorem 1 establishes how the incremental loss can be computed given the distribution of packets into timeslots, and the delay buffering capability at the switch. It follows that:

Corollary 1: $(D = 0) \Rightarrow (\delta = \rho)$.

Proof: Setting $D = 0$ in (1) yields $\delta = f^1(1)$. Also recall that $\rho = E[X(n)]/n$, which for any $X(n)$ with stationary increments equals $E[X(1)]/1 = P\{X(1) = 1\} = f^1(1)$. Thus $\delta = \rho$. □

This shows that if the switch has no FDL buffers, i.e., $D = 0$, δ depends only on the mean packet rate, and interestingly, is *invariant* to whether the packets are clumped together or spaced apart. In general, however, when $D > 0$, the distribution of packets into slots does affect δ. We illustrate this using three examples.

Example 1 (Block placement) This is one where every block of N successive slots carries a randomly placed block of $k = N\rho$ contiguous packets. The random location of the block ensures that the process $X(n)$ denoting the number of packets carried in n slots has stationary increments. From (1), the incremental loss is determined by the probability that a random set of $D+1$ contiguous slots all contain packets, which is $\max(0, (k-D)/N)$. Thus,

$$\delta_{block} = \max(0, \rho - D/N)$$

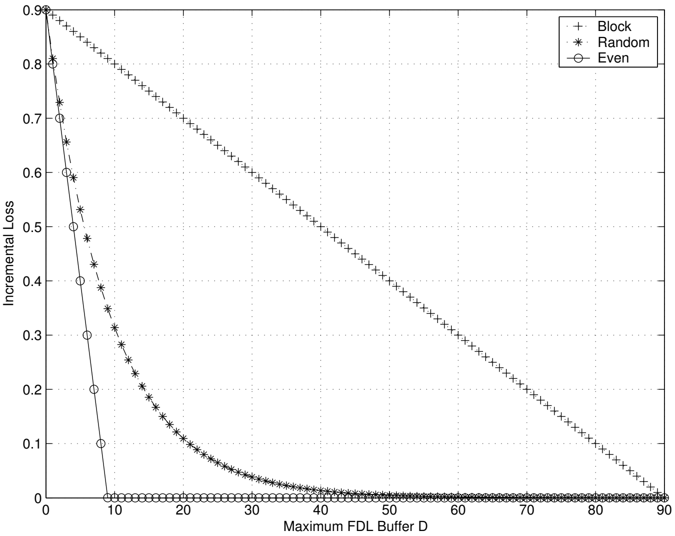

Fig. 2. Incremental Loss δ vs. delay-buffer size D for $\rho = 0.9$

Example 2 (Random placement) This is such that the packets are randomly and independently assigned to slots. A slot is thus occupied with probability ρ, and idle with probability $1 - \rho$. Stationarity is evident, and (1) yields the incremental loss to be the probability that any $D+1$ contiguous slots all carry slices, thus,

$$\delta_{random} = \rho^{D+1}$$

Example 3 (Even placement) Even placement is one where every window of $n = 1, 2, \ldots, \infty$ contiguous slots contains $\lfloor n\rho \rfloor$ or $\lceil n\rho \rceil$ packets. Thus $1 - P\{X(n) = \lfloor n\rho \rfloor\} = P\{X(n) = 1 + \lfloor n\rho \rfloor\} = n\rho - \lfloor n\rho \rfloor$. An arbitrary shift of an even placement yields another even placement, and stationarity holds. From (1), the incremental loss equals the probability that $D+1$ contiguous slots carry $D+1$ slices. For $\rho < 1$ we have $\lfloor (D+1)\rho \rfloor < D+1$, so the only way the $D+1$ slots carry $D+1$ slices is when $X(D+1) = \lfloor (D+1)\rho \rfloor + 1 = D+1$, i.e., when $\lfloor (D+1)\rho \rfloor = D$, which holds only when $D \leq \lfloor \rho/(1-\rho) \rfloor$. Further, $X(D+1) = D+1$ happens with probability $(D+1)\rho - \lfloor (D+1)\rho \rfloor$, which, from above, equals $(D+1)\rho - D$. Thus

$$\delta_{even} = \max(0, \rho - D(1-\rho)) \tag{2}$$

Figure 2 plots the incremental loss δ at the link as a function of switch delay-buffer capacity D for the three placement schemes discussed above, with average link utilization ρ fixed at 90%. All three cases exhibit identical losses at $D = 0$ (as predicted by corollary 1), but as D increases, the benefits of even placement are apparent: at $D = 9$ even placement realizes zero loss, while the block and random placements yield losses of about 30% and 90% respectively. We now show formally that among all placement schemes, even placement realizes minimum incremental losses.

Theorem 2: For given link utilisation ρ, the even placement of packets on the link minimises the incremental loss δ on that link.

Proof: We first derive a lower bound on δ as follows: consider a window of N contiguous slots, where $N \to \infty$. Of these N

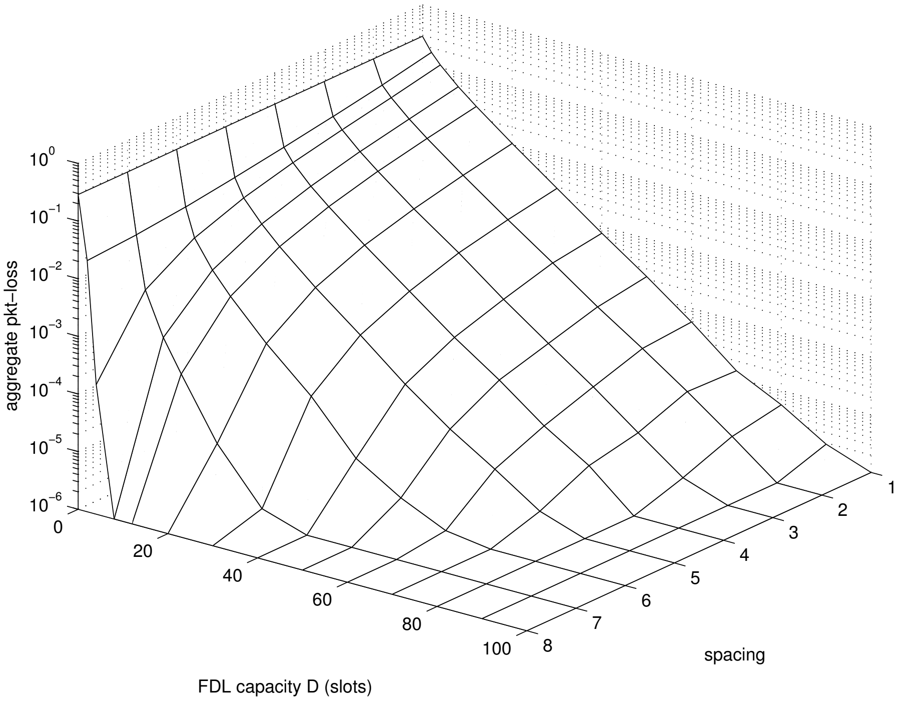

Fig. 3. Single-switch with Markovian traffic: aggregate packet loss as a function of packet spacing and FDL capacity D

slots, $k = \rho N$ are busy, i.e., carry packets, while the remaining $N - k$ are idle. Further, let b_i ($1 \leq i \leq m$) denote the size of the i-th contiguous block of packets, and g_i ($1 \leq i \leq m$) denote the size of the block of contiguous idle slots immediately following b_i. Then

$$\begin{aligned} b_1 + b_2 + \ldots + b_m &= k && (3) \\ g_1 + g_2 + \ldots + g_m &= N - k && (4) \\ \forall i,\ \ 1 \leq i \leq m: &\quad b_i \geq 1,\ \ g_i \geq 1 && (5) \end{aligned}$$

Constraint (5), in conjunction with (3) and (4), implies that

$$m \leq \min(k, N - k) \tag{6}$$

To compute δ, consider a packet arriving at the switch at random slot $s \in [1, N]$. If slot s falls on a gap, or on the j-th packet of block b_i where $j + D > b_i$, the incoming packet is schedulable. If, however, s falls on the j-th slot of block b_i where $j + D \leq b_i$, the packet is unschedulable. The loss probability is thus

$$\begin{aligned} \delta &= \frac{1}{N}\sum_{i=1}^{m}\max(0, b_i - D) \\ &\geq \frac{1}{N}\sum_{i=1}^{m}(b_i - D) \\ &= (k - mD)/N \quad \text{[from (3)]} \\ &\geq (k - (N-k)D)/N \quad \text{[from (6)]} \\ &= \rho - D(1 - \rho) \end{aligned}$$

This establishes a lower bound $\delta \geq \rho - D(1-\rho)$. Since δ_{even} in (2) achieves this lower bound, we have shown that even placement minimizes incremental losses. □

Theorem 2 establishes that even placement of packets on a link minimises the probability that an additional packet will be lost. This suggests that overall losses in the network can be minimised if traffic on *every* link in the OPS network is evenly spaced. This can be achieved by ensuring that:

- edge nodes evenly space the transmission of packets into the OPS network, and
- core nodes maintain this spacing when multiplexing traffic streams within the OPS network.

The former is achieved by having the edge nodes use their electronic buffers to condition the traffic before injection into the OPS network. The preservation of packet spacing within the optical core, however, is challenging because (a) the limited FDL buffering in the core switches may be insufficient to introduce the necessary gaps between packets when multiplexing streams, and (b) the complexity of the scheduling algorithms required for even spacing may be too high for practical implementation. For these reasons, we consider only the conditioning of traffic at the ingress, and defer the study of core node mechanisms for preserving packet spacing to future work. We show by simulation that edge traffic conditioning can by itself significantly reduce contention losses within the OPS network. The price paid for ingress conditioning is an increase in delay at the edge nodes. This tradeoff between loss and delay is studied via simulation next.

IV. Single-Switch Simulation Study

To understand the impact of edge traffic conditioning on the losses and delays in the OPS network, we consider two scenarios – the first is a minimal network setting demonstrating the effectiveness of ingress traffic conditioning, and consists of a single core-node connected to multiple edge switches, each fed with short-range dependent traffic. The second scenario, considered in the next section, uses a real-world network topology with realistic long-range dependent traffic.

All simulations utilise a single wavelength, since the absence of wavelength converters allows the wavelengths to be studied independently. Further, traffic flows are considered in only one direction. All links operate at 10Gbps. Dimensioning the optical slot size appropriately is important – larger slot sizes are associated with higher packet aggregation delays and/or inefficient filling of the slots, while smaller slot sizes may impose significant segmentation/reassembly overheads. We choose a slot size of $1\mu s$, which, at a bandwidth per wavelength of 10Gbps, carries an optical packet of size 1250 bytes. This slot size is commensurate with studies in the literature [26], and is also consistent with current optical crossbar technology (solid-state crossbars that can reconfigure within 20 ns have been demonstrated recently [27]).

Our scenario in this section consists of a single core switch, with 8 input lines, each connected to an ingress edge switch, and 1 output line, connected to an egress edge switch. Each ingress edge switch is offered short-range dependent traffic of fixed-size packets. More specifically, the traffic is a two-state Markov modulated fluid process (MMFP): in the "on" state, fluid traffic is generated at link rate (10Gbps), while in the "off" state no traffic is produced. The holding times in the on and off states are exponentially distributed, with mean holding time of $2\mu s$ in the on state, and average traffic rate of 1Gbps. This model represents bursty traffic that is short-range dependent. The fluid traffic is packetized into fixed-length packets, which is fed to the edge node. The edge switch in turn aggregates multiple such packets into an optical packet of fixed length, and transmits it in an appropriate slot (as determined by the traffic conditioning) to the core switch for forwarding to the destination edge switch. The core switch output link in this scenario is loaded at 8 Gbps or $\rho = 0.8$ of link capacity. At the core switch, multiple arriving packets may contend for the output link. The core switch has FDLs that can delay packets up to D slots, and for a contending packet uses the shortest available FDL that resolves the contention; if no such FDL is available the packet is dropped.

We study the effect that edge traffic conditioning has on the contention losses at the core switch and on the end-to-end delays experienced by the packets. The traffic conditioner is parametrised by the minimum allowable spacing (in slots) between any two successive packets it releases into the optical network. Thus a spacing of 1 indicates that two packets can emerge back-to-back (i.e. 1 slot apart), and corresponds to no conditioning, whereas a larger spacing mandates idle slots between any two successive packet transmissions. Stated another way, the edge traffic conditioner that uses spacing s restricts the output rate to $1/s$ of link rate. Note that for a stable system, s has to be in the range $[1, 1/\rho_f)$, where ρ_f denotes the average traffic rate at the traffic conditioner. Each point in the simulation plots in this paper corresponds to a run of at least 40 million packets.

Figure 3 shows, for various FDL capacities (D) at the core switch, the effect of varying the spacing between successive packets at every edge node, on the total losses at the core switch egress link. Note first that when $D = 0$, the losses are invariant to traffic conditioning, as predicted by corollary 1. Now observe how losses fall as D increases, and also as the packet spacing increases. This shows that edge traffic conditioning can drastically reduce the amount of FDL buffering required at the core switch for a desired loss rate in the OPS network. For example, in the absence of conditioning (this corresponds to a minimum packet spacing of 1), losses of 10^{-4} require FDL buffers of length 50. In comparison, with edge traffic conditioning where the packet spacing is 8 (i.e. one in eight slots is available for transmission), FDL capacity 4 suffices at the core switch to realise such a loss rate. This translates to a significant reduction in cost, both in terms of the FDLs required and the crossbar size, which demonstrates that edge traffic conditioning can help reduce network cost considerably while delivering the desired loss performance.

flow#	src → dest	traffic type	hops
$F1$	RNSH → Nepean	medical	1
$F2$	Nepean → ARRC	management	2
$F3$	Riverside → Conservatorium	functions	1
$F4$	Conservatorium → UWA	music collab	2
$F5$	CSIRO-Marsfield → ARRC	intranet	3
$F6$	MQU → UNSW	university	1
$F7$	UNSW → UMel	university	1
$F8$	UMel → UWA	university	1

TABLE I
CeNTIE Network Flows Simulated

V. CeNTIE Network Simulation Study

We now study the impact of edge traffic conditioning in a real-world network topology with realistic traffic flows carrying long-range dependent traffic. Figure 4 shows part of the CeNTIE network, a trans-continental Australian research network [28] with MANs in Sydney, Canberra, Melbourne, and Perth. It includes end-user research groups from the health, education, film post-production, and finance industries. We simulate a subset of the CeNTIE network, with logical topology and fibre lengths shown in Figure 5. There are 4 core switches in the chosen topology – two in Sydney, at CSIRO-Marsfield and the University of Technology, Sydney (UTS), and one each in Melbourne and Perth. There are four edge switches connected to the Marsfield core switch - one each at CSIRO-Marsfield, CSIRO-Riverside, MacQuarie University (MQU), and the Royal North Shore Hospital (RNSH). The UTS core switch is connected to three edge switches – the Conservatorium of Music (Con), University of New South Wales (UNSW), and Nepean Hospital. At Melbourne, there is an edge switch at the University of Melbourne (UMel), while Perth has two edge switches, at the Australian Resource Research Centre (CSIRO-AARC) and the University of Western Australia (UWA). All the above mentioned sites are either currently live or going live soon on the CeNTIE network.

For the simulations, we selected eight traffic flows typical of the usage of the CeNTIE network. These flows are depicted in Figure 5, and Table I shows their characteristics including the type of traffic and the number of core-links traversed. The diversity in hop-lengths and end-to-end propagation delays give a representative sampling of traffic flows in the CeNTIE network. Flows $F1$-$F4$ carry time-critical traffic, and are thus not subjected to traffic conditioning at the ingress edge nodes. Flows

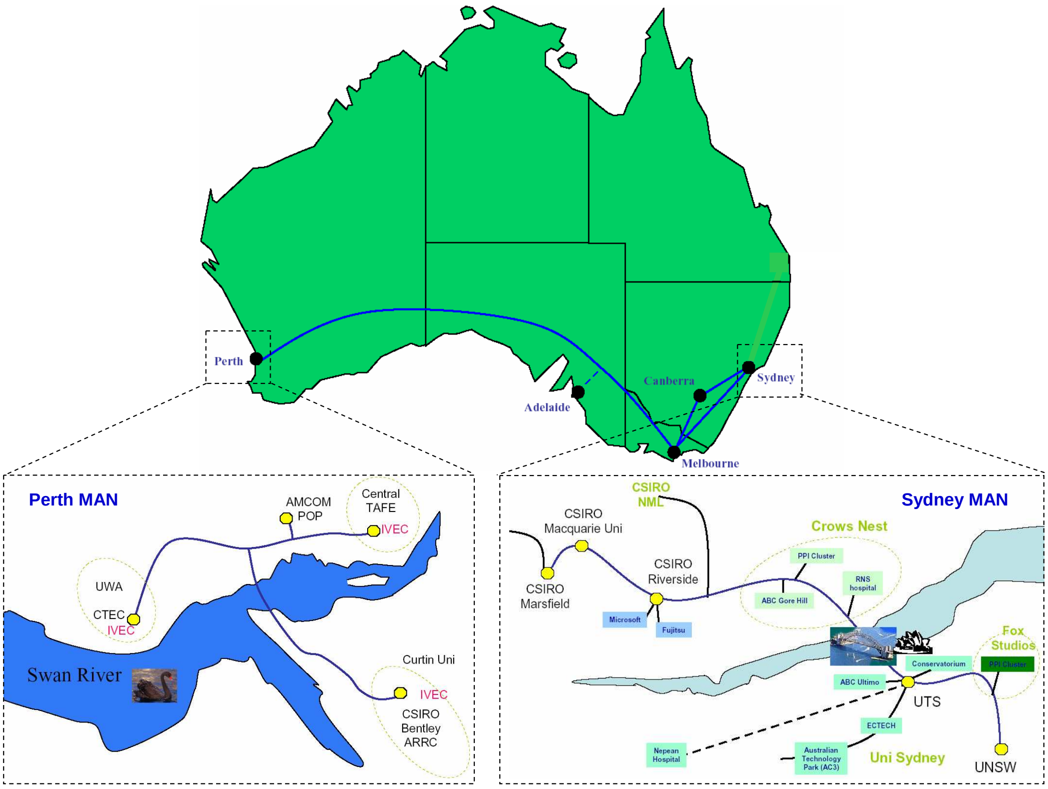

Fig. 4. CeNTIE Network

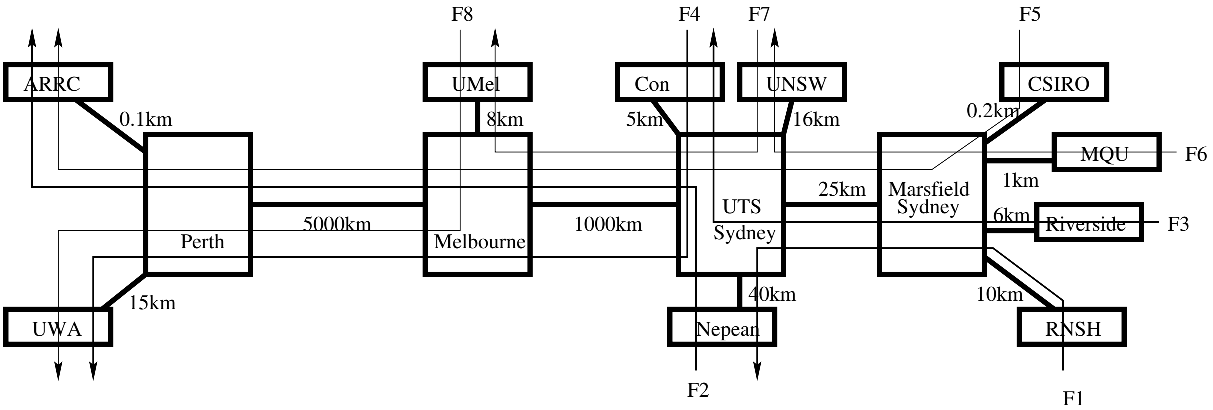

Fig. 5. CeNTIE Simulation Topology and Flows

$F5$-$F8$ carry intranet and university traffic, and are considered non-time-critical, and are subject to ingress traffic conditioning.

The slot-size and link capacity are the same as the previous scenario, namely $1\mu s$ and 10Gbps respectively. The simulation considers a single wavelength with unidirectional traffic flows. Each flow generates traffic at a mean rate of 1.5 Gbps, offering a load $\rho_f = 0.15$ of link capacity. Each of the three core links carries 4 flows, corresponding to a load of 6 Gbps or $\rho = 0.6$ on each core link. We believe such a loading scenario is realistic.

Traffic in each flow is long range dependent (LRD), and based on Norros' self-similar traffic model [29]. This model combines a constant mean arrival rate with fractional Gaussian noise (fGn) characterised by zero mean, variance σ^2 and Hurst parameter $H \in [1/2, 1)$. We use our filtering method [30], related to the FFT-based methods described in [31], [32], to generate, for a chosen H, a sequence $\{x_i\}$ of normalised fGn (zero mean and unit variance). A discretisation interval Δt is chosen, and each x_i then denotes the amount of traffic, in addition to the constant rate stream, that arrives in the i-th interval. Specifically, the traffic y_i (in bits) arriving in the i-th interval of length Δt seconds is computed using:

$$y_i = max\{0, \rho_c \Delta t + s x_i\} \tag{7}$$

where ρ_c denotes in bits-per-second the constant rate stream,

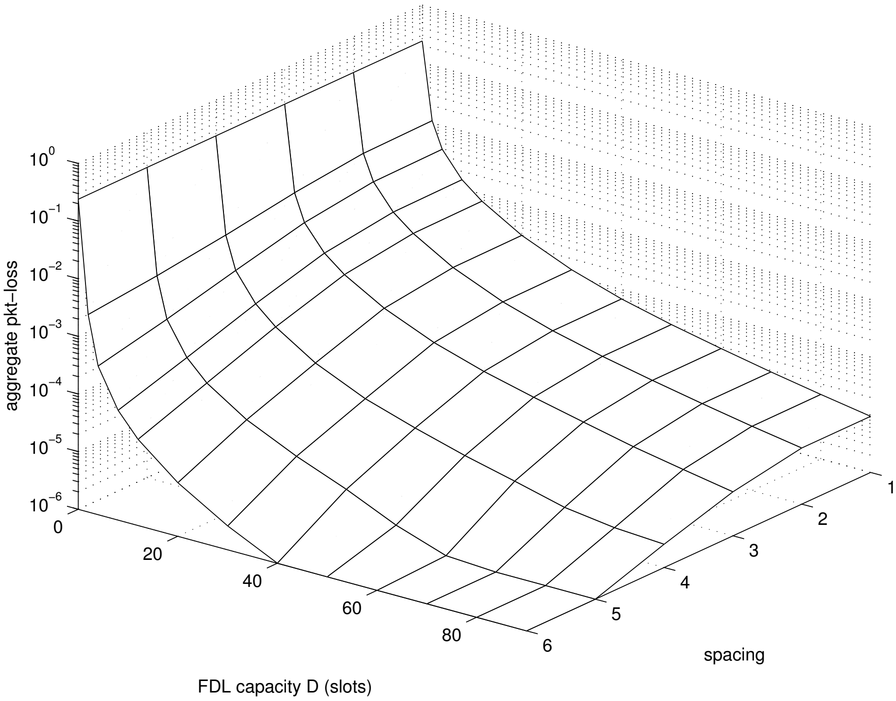

Fig. 6. CeNTIE network with LRD traffic: aggregate packet loss as a function of packet spacing and FDL capacity D

and s is a scaling factor. The scaling factor s determines the variable component of the traffic arrival rate, and ρ_c is selected to achieve the desired mean arrival rate ρ_f for each flow. For this work we set the Hurst parameter at $H = 0.85$ and the discretisation interval $\Delta t = 1.0\mu s$. The burstiness at short time-scales is set to a relatively high level by choosing $\rho_c \Delta t/s = 0.5$, and ρ_c is then adjusted to give the desired mean traffic rate of 1.5Gbps. This fluid traffic is then packetised into fixed-length packets before being fed to the edge nodes.

The edge nodes assemble incoming packets into optical packets of fixed length (1250 bytes). For flows $F1$-$F4$, the assembled optical packets are transmitted in FIFO order at the earliest available slot, since these flows carry time-critical traffic. The transmission of optical packets for flows $F5$-$F8$ by the appropriate edge into the OPS network, however, is paced out according to a parameter, namely the minimum spacing between successive packets. Figure 6 shows the aggregate packet loss probability (on log scale) in the network as the spacing for flows $F5$-$F8$, and the FDL capacity D (in slots) at each core switch vary. Again, the losses fall as D increases, and also as spacing increases, again confirming that edge traffic conditioning does reduce losses in the optical core. It is interesting to compare the aggregate loss plots of Figures 3 and 6 and note that in the absence of ingress traffic conditioning (i.e. when $\phi = 0$), the losses fall off exponentially for SRD (Markovian) traffic and sub-exponentially (i.e., as a power-law) for LRD traffic as the FDL buffer D at the core switches is increased. This form of the loss curve is expected for LRD traffic, and signifies that the incremental cost of FDLs required to reduce the loss by a desired amount gets progressively higher. This reinforces our proposition that in lieu of employing more FDLs, ingress conditioning can reduce contention losses cost-effectively.

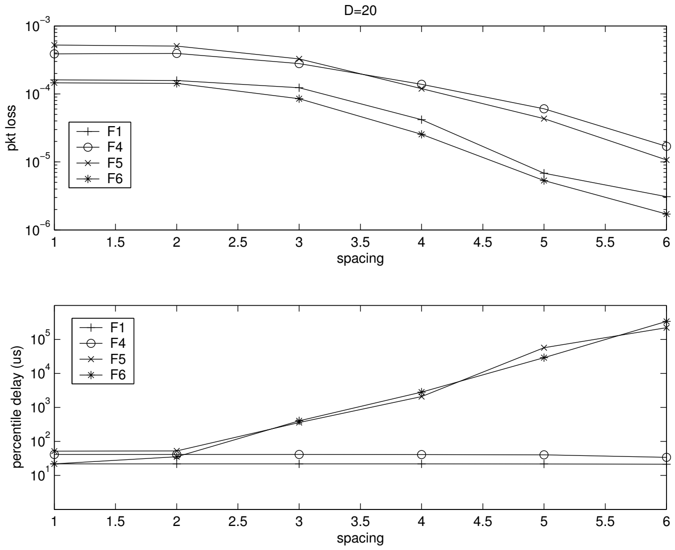

Fig. 7. Per-flow loss and delay for LRD Traffic, $D = 20$ slots

Ingress traffic conditioning reduces contention losses, but at the ingress shaper introduces queueing delay which increases with packet spacing. We evaluate the resulting per-flow loss-delay tradeoff. For each flow, we measure the packet loss and the tail of the delay disrtibution (rather than the mean). The tail is measured by the 99.999-percentile end-to-end packet delay (excluding propagation and packet aggregation delays), i.e., no more than 1 in 10^5 packets suffer greater delay.

Figure 7 plots, for $D = 20$, the loss and delay as the ingress packet spacing for flows $F5$-$F8$ is varied. For clarity, only four flows are depicted: $F1$ and $F4$ that respresent time-critical (hence unspaced) flows of hop-length 1 and 2 respectively, and flows $F5$ and $F6$ that are subjected to traffic conditioning and have hop-lengths 3 and 1 respectively. Note that when no flow is conditioned (spacing is 1), losses are higher for longer flows – $F5$ being the longest with 3 hops, followed by $F4$ which is 2 hops, followed by $F1$ and $F6$ that are 1 hop each. However, as the inter-packet spacing for non-time-critical flows increases, losses go down for all flows, but more so for the flows that are conditioned ($F5$ and $F6$ depicted in this case), though the difference is not large. For a spacing of 5, all flows experience a reduction of a order of magnitude in their losses.

The end-to-end delays are also depicted in the same figure, as the spacing for flows $F5$-$F8$ is varied. As expected, the delays of flows $F1$-$F4$ are invariant to the spacing of the other flows, and hence $F5$ and $F6$ appear as horizontal lines in the delay plot. The delays for $F5$ and $F6$ rise as the spacing increases. At a spacing of 5, the delays are of the order of a few tens of milliseconds, which may well be acceptable for non-time-critical traffic especially if it helps reduce losses in the network by more than an order of magnitude. The spacing parameter thus allows delay to be traded off against loss. By choosing an appropriate spacing metric for the input traffic conditioners of non-time-critical traffic, the network operator can select a desired loss-delay operating point for the flows in the network.

VI. Conclusions and Future Work

We considered a slotted OPS network with fixed size packets, and studied the effect of ingress traffic conditioning, namely controlling the spacing of optical packets transmitted into the OPS network. We demonstrated analytically that the even spacing of packets minimises incremental losses caused by contention on the optical link. We justified the efficacy of ingress traffic conditioning in a simple single core-switch topology carrying short-range dependent traffic streams. We then quantified the impact of the packet spacing of non-time-critical traffic flows in a real Australian network topology carrying representative traffic flows with long-range dependent traffic patterns. We showed that edge conditioning is universally effective in reducing contention losses, enabling OPS networks with acceptable losses to be realised using smaller amounts of FDLs. A reduction in loss comes at the cost of increased end-to-end delay for the conditioned flows, and the trade-off can be selected by the network operator. Our proposal of edge traffic conditioning therefore allows the network operator to choose a desired loss-delay tradeoff for flows in the OPS network.

This paper has only studied edge traffic conditioning; the preservation of packet spacing *within* the OPS network by means of intelligent core node scheduling algorithms is being investigated by our current research. We are also working on extending our study to include a more thorough study of multiple QoS classes in the network.

References

[1] R. Ramaswami and K. N. Sivarajan. *Optical Networks, A Practical Perspective*. Morgan Kaufmann, second edition, 2002.

[2] I. Chlamtac, A. Farago, and T. Zhang. Lightpath (wavelength) routing in large WDM networks. *IEEE J. Selected Areas in Comm.*, 14(5):909–913, Jun 1994.

[3] B. Mukherjee, D. Banerjee, S. Ramamurthy, and A. Mukherjee. Some Principles for Designing a Wide-Area Optical WDM Network. *IEEE/ACM Trans. Networking*, 4(5):124–129, Oct 1996.

[4] C. Qiao and M. Yoo. Optical Burst Switching (OBS) – A New Paradigm for an Optical Internet. *J. of High Speed Networks*, 8(1):69–84, 1999.

[5] I. Chlamtac, V. Elek, A. Fumagalli, and C. Szabo. Scalable WDM access network architecture based on photonic slot routing. *IEEE/ACM Trans. Networking*, 7(1):1–9, Feb 1999.

[6] S. Yao, S. Dixit, and B. Mukherjee. Advances in photonic packet switching: an overview. *IEEE Comm. Magazine*, 38(2):84–94, Feb 2000.

[7] A. Carena et al. OPERA: An Optical Packet Experimental Routing Architecture with Label Swapping Capability. *J. Lightwave Tech.*, 16(12):2135–2145, Dec 1998.

[8] C. Guillemot et al. Transparent Optical Packet Switching: The European ACTS KEOPS Project Approach. *J. Lightwave Tech.*, 16(12):2117–2134, Dec 1998.

[9] D. Hunter et al. WASPNET: A Wavelength Switched Packet Network. *IEEE Comm. Magazine*, 37(3):120–129, Mar 1999.

[10] D. Wonglumson et al. HORNET: A packet switched WDM network: Optical packet transmission and recovery. *IEEE Photonics Tech. Letters*, 11(12):1692–1694, Dec 1999.

[11] L. Dittmann et al. The European IST Project DAVID: A Viable Approach Toward Optical Packet Switching. *IEEE J. Selected Areas in Comm.*, 21(7):1026–1040, Sep 2003.

[12] D. Hunter, M. Chia, and I. Andonovic. Buffering in Optical Packet Switches. *J. Lightwave Tech.*, 16(12):2081–2094, Dec 1998.

[13] S. L. Danielsen, P. B. Hansen, and K. E. Stubkjaer. Wavelength conversion in optical packet switching. *J. Lightwave Tech.*, 16(12):2095–2108, Dec 1998.

[14] V. Eramo and M. Listani. Packet loss in a bufferless optical WDM switch employing shared tunable wavelength converters. *J. Lightwave Tech.*, 18(12):1818–1833, Dec 2000.

[15] F. Forghierri, A. Bononi, and P. R. Prucnal. Analysis and comparison of hot-potato and single-buffer deflection routing in very high bit rate optical mesh networks. *IEEE Trans. Commun.*, 43(1):88–98, Jan 1995.

[16] S. Yao, B. Mukherjee, S. J. B. Yoo, and S. Dixit. A unified study of contention-resolution schemes in optical packet-switched networks. *J. Lightwave Tech.*, 21(3):672–683, Mar 2003.

[17] G. N. Rouskas and L. Xu. Optical packet switching. In *Optical WDM Networks: Past Lessons and Path Ahead*, chapter 1. Kluwer, 2003.

[18] M. J. Karol. A shared-memory optical packet (ATM)switch. In *Proc. 6th IEEE Wksp. Local and Metro Area Networks*, pages 205–211, 1993.

[19] J. Ramamirtham and J. Turner. Time Sliced Optical Burst Switching. In *Proceedings of INFOCOM 2003*, San Francisco, CA, Mar 2003.

[20] S. Yao, F. Xue, B. Mukherjee, S. J. B. Yoo, and S. Dixit. Electrical ingress buffering and traffic aggregation for optical packet switching and their effect on tcp-level performance in optical mesh networks. *IEEE Comm. Magazine*, 40(9):66–72, Sep 2002.

[21] A. Ge, F. Callegati, and L. S. Tamil. On optical burst switching and self-similar traffic. *IEEE Comm. Letters*, 4(3):98–100, Mar 2000.

[22] F. Xue and S. J. B. Yoo. Self-similar traffic shaping at the edge router in optical packet-switched networks. In *Proc. IEEE ICC 2002*, pages 2449–2453, New York, NY, Apr 2002.

[23] R. Nejabati and D. Simeonidou. Class-based aggregation in optical packet switched WDM networks. In *TERENA Netw. Conf.*, pages 19–22, Zagreb, Croatia, May 2003.

[24] F. Xue, S. Yao, B. Mukherjee, and S. J. B. Yoo. The performance improvement in optical packet-switched networks by traffic shaping of self-similar traffic. In *Proc. OFC 2002*, Anaheim, CA, Mar 2002.

[25] G. Hébuterne and H. Castel. Packet aggregation in all-optical networks. In *Proc. ICOCN 2002*, Singapore, Nov 2002.

[26] D. Careglio, J. Pareta, and S. Spadaro. Optical slot dimensioning in IP/MPLS over OPS networks. In *Proc. WOAN 2003.*, Zagreb, Croatia, Jun 2003.

[27] T. McDermott and T. Brewer. Large-Scale IP Router using a High-Speed Optical Switch Element. *J. Optical Networking*, 2(7):229–240, Jul 2003.

[28] Terry Percival. An Introduction to CeNTIE. Presentation. *http://www.centie.org/docs/CeNTIE-web-intro.ppt*.

[29] I. Norros. On the use of Fractional Brownian Motion in the Theory of Connectionless Traffic. *IEEE J. Selected Areas in Comm.*, 13(6):953–962, Aug 1995.

[30] D. Ostry. Fast synthesis of accurate fractional Gaussian noise. *Submitted for publication*, 2004.

[31] A. T. A. Wood and G. Chan. Simulation of Stationary Gaussian Processes in $[0,1]^d$. *J. Computational and Graphical Statistics*, 3(4):409–432, Dec 1994.

[32] C. R. Dietrich and G. N. Newsam. Fast and Exact Simulation of Stationary Gaussian Processes through Circulant Embedding of the Covariance Matrix. *SIAM J. Scientific Computing*, 18(4):1088–1107, July 1997.