Statistical multiplexing in optical flow switching networks

Kai Wu[a] and Qingji Zeng[b*]

[a] Univ. of Nebraska-Lincoln, NE, USA

(Previously Shanghai Optical Networking Technology Co., Ltd. China)

[b] Shanghai Jiao Tong University, Shanghai, China

ABSTRACT

In IP over WDM networks, optical paths (OP's) form a virtual topology seen from the upper IP layer, and greatly relieve the burden of electronic processing thanks to optical bypasses. However, in contrast to the circuit-switching, connection-oriented and coarser OP's, the IP is packet-switching, connectionless and finely granular in nature. Special care should be taken in the control plane to bridge the gap. In our research, we incorporate the idea of optical flow. Optical flows are optically switched in intermediate nodes and converted into electronic form at termination to treat small nested flows. Two modes of optical flow switching (OFS) are examined. One is optical burst switching (OBS), which sends control packet ahead of data to announce a switching and make future reservations. The other is Optical Grooming Switching (OGS), which puts packet flows into finer container with discrete yet flexible bit rate. We find that both of these two modes use statistical multiplexing to bridge the gap between IP packets and WDM OP's, and such function is fulfilled respectively by scheduling in OBS and by framing in OGS. Taking snapshots of statistical multiplexing, we formulate the bandwidth distribution (BWD) of the whole OFS networks and the bandwidth usage (BWU) over each node pair. The BWD can be expressed by the product of a traffic-independent grooming matrix with the traffic demand matrix under a given virtual topology. Changes to the grooming matrix are reflected by adjustments in framing or scheduling. Three distinct states of BWU are also identified to determine the initiation of such changes. The BWU states are hence proposed to add to the extension of existing routing protocols.

Keywords: optical flow, optical burst switching, grooming, statistical multiplexing, IP over WDM

1. INTRODUCTION

One of the most important issues in integrating WDM and IP networks is to define the interface between them. Chiefly various definitions could be classified into three categories: a) optical path; b) optical flow; c) optical packet.

In the first category, the WDM layer operates in a circuit-switching mode. Optical paths (OP) could bypass intermediate nodes without opto-electronic conversion[1]. These OP's form a virtual topology seen from the upper IP layer, and greatly relieve the electronic processing burden thank to the optical bypasses. However, since in contrast to the connection-oriented and coarser OP's, the IP is packet-switching, connectionless and finely granular in nature, a special (may be complex) adaptation layer should be implemented in between to bridge the gap.

In the third category, the IP packets could take the optical form directly. Packet header processing and forwarding are handled in the optical domain[2]. However such architectures are not achievable without mature technology in optical buffer and processing, which is embryonic compared to the electronic counterpart.

To walk out the dilemma, a third way is proposed to incorporate the idea of optical flow (the second category). A flow is an aggregation of streams of packet data with common routes and service parameters. It features a certain aggregated bandwidth and duration and can be recursively defined. Two modes of optical flow switching (OFS) are examined. One is optical burst switching (OBS), which sends control packet ahead of data to announce a switching and make future reservations. The other is Optical Grooming Switching (OGS), which put packet flows into finer container with discrete yet flexible bit rate.

In this paper, we examine in deep these two models, and identify where and how statistical multiplexing works to enhance the performance of integrating IP packet data and WDM optical paths. Section 2 and 3 will deal OBS and OGS respectively. Section 4 then makes an analysis on the general OFS model, while section 5 draws conclusion.

---

[*] Further author information: (Send correspondence to Kai Wu)
Kai Wu: E-mail: kwu@cse.unl.edu

Optical Networking, Chunming Qiao, Shizhong Xie, Editors, Proceedings of SPIE Vol. 4585,
APOC 2001, Beijing, China (2001) © 2001 SPIE · 0277-786X/01/$15.00

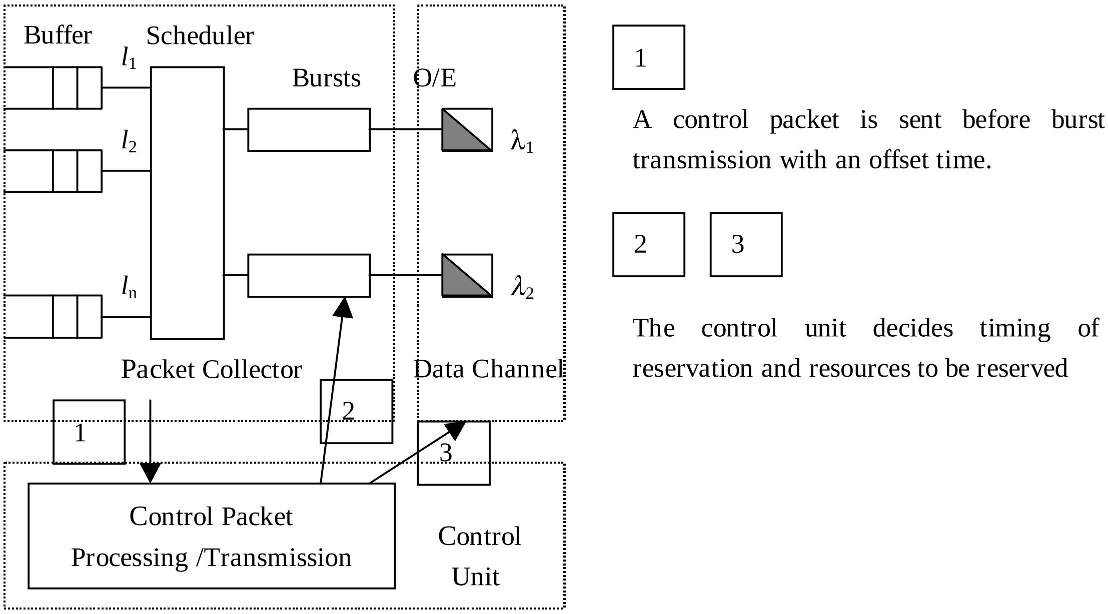
2. MODEL OF OPTICAL BURST SWITCHING NETWORK

The first approach we considered is Optical Burst Switching, or OBS. OBS takes advantage of the buffering and processing capabilities of electronics at network edge, combined with benefit of optical bypass and wavelength routing in the core.

There are several schemes in implementing OBS, such as TAG, JET, IBT, RFD, JIT, Horizon[3,4,5]. These approaches mainly differ in time to initiate allocation and de-allocation of resources; however, they share the same main features listed as following:

1) **Burst switching**. The optical switches switch on a burst-by-burst basis and transmit bursts at the full speed of a wavelength channel. A switch maintains the state of the established wavelength channel and occupies the acclaimed resources during the lifetime of a burst, and after burst transmission is over, it releases the resources.
2) **Optical bypass**. Data packets are packaged into bursts at the network edge and then transmitted through intermediate nodes without or with minimum optical-electrical conversion. Thus in the core, buffer and processing requirement may be eliminated.
3) **Separated processing and transmission of control packets**. The control information is sent on a separate band (on different wavelength or sub-carrier multiplexed channel), and before transmission of the burst data with an offset time. It is also processed in intermediate nodes without re -packaging of the burst payload data
4) **Reservation mechanism**. Reservation is used to efficiently utilize the bandwidth of the established wavelength channel and enhance sharing of resources. Through some improvements to the reservation schemes, the OBS can be capable of QoS support.

Based on these four observations, we identify three key components of OBS edge router, as is shown on Fig-1.



A control packet is sent before burst transmission with an offset time.

The control unit decides timing of reservation and resources to be reserved

Figure 1. Three key components of an OBS edge router

The key components are:

1) **Packet collector**. Packet collector collects the input packets and places them in certain burst boxes to form bursts. The incoming buffer acts as an assembler for data flows. The bit rate, collected and estimated length, priority and routing of the flows are inputs to the scheduling algorithm within the scheduler to determine the sequence in which the flows are placed into free burst boxes. A free burst box is the burst

input to a free data channel. The scheduler also makes burst request to control unit to initiate a control infomation packet.

2) **Data channel**. Data channel is the abstraction for a potential wavelength channel to transmit a burst. There are two states for a data channel. In the reservation state, resources such as wavelength, transceiver and OXC ports are assigned to the data channel, and hence a wavelength channel is setup. After the burst is transmitted, the wavelength channel is torn down, and the resources are released. In the release state, the data channel just waits for new reservation command from control unit.

3) **Control unit**. The control unit takes care of the control packet. The control packet should include enough information for burst routing, offset time (on which the burst will arrive) and burst length (the length known plus an estimated length the buffer assembled during the offset time).

Note in the packet collector, if grooming is allowed, several lower-rate flows sharing the same routing could be multiplexed to form a higher-rate flow and placed together into a burst box; otherwise, these flows could occupy the burst box in turn to minimize re-configuration of wavelength channels. Either case, statistical multiplexing is introduced to increase utilization of bursts or data channels.

It is important to examine some time relations among these three units.

For a data channel to transmit a burst four parts compose a reservation of a data channel, i.e., the time for a reservation state is at least

$$T_{\text{reservation (Min)}} = T_{\text{setup}} + T_{\text{burst}} + T_{\text{prop}} + T_{\text{teardown}} \qquad (1)$$

Where T_{setup} and T_{teardown} are time for physical actions, if any, respectively to setup and to teardown of a wavelength channel after receiving the allocation or de-allocation request from the control unit. The T_{prop} is the propagation time the burst takes to pass the channel. It is added since teardown could not happen before the last bit of the burst is transmitted. And last T_{burst} is the transmission time for the burst at the full speed of the wavelength channel. The three parts except for the T_{burst} are a cost incurred by re-configuration. If re-configuration takes place frequently, these parts could be accumulated to be a considerate part, and lessen the bandwidth utilization. Therefore, statistical multiplexing (either with or with out grooming) is strongly recommended to reduce the chance of re-configuration, in which case the time for setup and teardown is zero, and even T_{prop} could be excluded in computing the minimum $T_{\text{reservation}}$.

We define the re-configuration bandwidth cost as

$$C_{reconfig} = \frac{T_{reservation} - T_{burst}}{T_{reservation}} \qquad (2)$$

and bandwidth utilization as

$$U = \frac{T_{burst}}{T_{reservation}} \qquad (3)$$

Obviously, we have,

$$C_{reconfig} = 1 - U \qquad (4)$$

A burst data will not be transmitted before the resource reservation of the previous burst occupied is released. Thus an average end-to-end delay for a burst is

$$\overline{T}_{delay} = \overline{T}_{reservation} + \overline{T}_{release} \qquad (5)$$

where T_{release} is the release time or idle time for a data channel, in equation (5) the average is made over all used data channels.

Again, in the case of statistical multiplexing of flows sharing same routing, re-configuration is avoided and delay is identical to all these flows.

3. MODEL OF OPTICAL GROOMING SWITCHING NETWORK

Grooming has been widely studied in SDH/SONET over WDM networks. The aim is to reduce the number of ADM devices and/or wavelength. Both static and arbitrary traffic models have been examined. With finer granularity and faster switching speed, combined with the benefits to form VWP (wavelength conversion) these ADMs could be used simply as a grooming switch to bridge the gap between IP packets and WDM optical flows. To maximize optical bypass, in the core, sparse or limited grooming switching can be deployed.

Under this network model, packets are colleted to form sub-wavelength flows with a variable bit rate, reaching one or more multiples of basic line speed. The basic line speed is the switching element the grooming switches employ. Similar to OBS in section 2, we can identify three key components for an OGS edge router, as shown in Fig-2

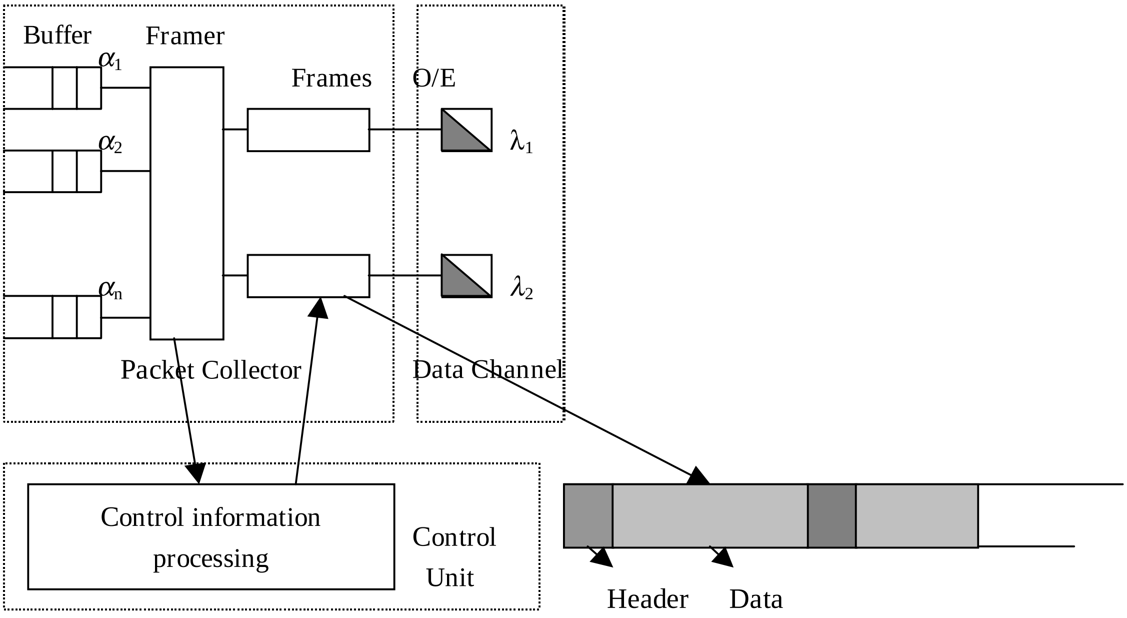


Figure 2. Three key components of an OGS edge router

Again, there are three key components: packet collector, data channel and control unit. Control information could be read out from the header, and processed separately without re-packaging of the payload data. The major difference from OBS is the framer rather than scheduler in the packet collector of OGS. The framer employs statistical multiplexing directly. The framing is assumed to aggregate flows of various rates (α_i in Fig 2) into one frame. Inside the frame, length of each data channel is proportional to their bit-rates, or adjusted for some QoS considerations. The header includes the label and basic data features like bit-rate, priority, TTL etc.

4. ANALYSIS OF GENERAL OPTICAL FLOW SWITCHING

In a certain period, the bursts in OBS could be taken as equivalent to data flows in OGS. When statistical multiplexing employed, bursts are in fact flows transmitted sequentially at full speed of a wavelength. Such equivalence is analogous to what serial bus to parallel bus. Therefore in finding equations for bandwidth usage, we consider flow framing in OGS only. For OBS similar equations could be found.

Taking a snapshot of the core OGS network, we find a set of optical flows. This is not much unlike the virtual topology in wavelength-routed networks. In our formulations we care only the bandwidth of data links the optical flow implements, and denote them as matrix **d**, where the index is node pair **s**. Then take a snapshot for the whole OGS including the edge, and we have a set of flows with defined bandwidth. We consider only the flows with different routes, and denote the flow bandwidth by matrix α, indexed by different route **r**. Here route is a concatenation of data link rather than a real edge, though the routing protocol in MPLS is one layered. We may obtain the route by combining the routing results of flows and optical flow at these two stages.

Further, we notice that the combined bandwidth of input could be obtained at the output of incoming buffers. We denote the traffic demand as matrix **t**, indexed by node pair **s**. Finally, the matrix α could further be split into two matrixes, α_1 and α_m, which denote bandwidth respectively for one-hop (taking one data link) traffic and combined multi-hop (taking more than one data link) traffic. Both of them are indexed by node pair **s**.

We have bandwidth constraints over each node pair and data link.

$$\mathbf{E}\cdot\boldsymbol{\alpha}^{(1)} + \mathbf{E}\cdot\boldsymbol{\alpha}^{m} \geq \mathbf{t} \tag{7}$$

$$\mathbf{E}\cdot\boldsymbol{\alpha}^{(1)} + \mathbf{C}\cdot\boldsymbol{\alpha}^{m} \leq \mathbf{d} \tag{8}$$

Here **E** denotes an identity matrix, **C** refers to bandwidth distribution of multi-hop traffic.

Let's define some derived variables:

$$\Delta = \mathbf{d} - \mathbf{t} \tag{9}$$

Δ refers to the difference between provisioned bandwidth and the demanded bandwidth over each node pair.

$$\mathbf{S} = \mathbf{d} - \boldsymbol{\alpha}^{(1)} \geq 0 \tag{10}$$

S refers to the spare capacity left for grooming of multi-hop traffic.

$$\mathbf{R} = \mathbf{t} - \boldsymbol{\alpha}^{(1)} \geq 0 \tag{11}$$

R refers to the residue traffic other than one-hop traffic; this traffic must be multi-hop routed.

The non-negativeness of **R** and **S** sets an upper bound for one-hop traffic α_1. Finally we define

$$\mathbf{G} = \mathbf{C}\cdot\boldsymbol{\alpha}^{m} \tag{12}$$

G refers to the bandwidth groomed over each node pair.

From the meaning of **Δ**, **S**, **R** and **G**, we also have

$$\Delta = \mathbf{S} - \mathbf{R} \tag{13}$$

$$\mathbf{R} \geq \boldsymbol{\alpha}^{m} \tag{14}$$

$$\mathbf{G} = \mathbf{C}\cdot\boldsymbol{\alpha}^{m} \leq \mathbf{S} \tag{15}$$

From these equations we can determine three distinguished states of connections over each node pair, as shown in the following table. These states are determined by bandwidth difference **Δ** and above equations.

Table 1 Bandwidth usage states for connections over one node pair

State	I	II	III
Bandwidth difference	$\Delta >> 0$	$\Delta \approx 0$	$\Delta << 0$
One-hop traffic carried	$\alpha^{(1)} > t$	$\alpha^{(1)} = d - S$	$\alpha^{(1)} = d$
Multi-hop traffic groomed	$\alpha^{m} = 0$	$\alpha^{m} > R$	$\alpha^{m} > R$
Residue traffic for multi-hop routed	$R = 0$	$R = S - \Delta$	$R = -\Delta$
Spare space for traffic grooming	$S = \Delta$	$S > 0$	$S = 0$

State I holds when one-hop traffic is dominant. In this state, there may be plenty of space for grooming and also a high risk of bandwidth wastage. This is the case when aggregated bandwidth for flows is far less than the wavelength bandwidth in OGS, or there's a big interleave between bursts in OBS.

State II holds when there's a non-zero residue multi-hop traffic, and there's also some space for grooming. It's a tradeoff between one-hop and multi-hop traffic, since the higher the one-hop, the lower the multi-hop traffic together with the lower room for grooming.

State III holds when multi-hop traffic is dominant and there's no room for grooming. In this state, traffic runs the risk of more delay and buffer overflow.

At any point of network operation, as long as we could find the grooming matrix **C** satisfying upper equations, the network is in a good state where each node pair remains in one of the three states, and there is no need to initiate rerouting of flows. However as we can see, state I is risky in wasting bandwidth while state III is apt to causing delay, so some traffic engineering operations should be undertaken under our framework to reduce states I and III. And especially, special algorithms in framing or scheduling may improve the whole network.

5. SUMMARY

We have discussed two approaches for optical flow switching: Optical Burst Switching and Optical Grooming Switching. By summarizing the similarities of the two, we've identified three key components in general Optical Flow Switching networks. Statistical multiplexing is fulfilled respectively by scheduling in OBS and by framing in OGS in the packet collector unit. By leveraging statistical multiplexing, re-configuration is minimized and bandwidth utilization is enhanced. An analysis is made under this general model, using formulations regulating the bandwidth usage and distributions, which provides a deeper understanding of bandwidth usage over the whole network and over each node pair as well. The results can be used to evaluate the design of a routing and signaling under optical flow network, and algorithms in framing and scheduling, which belong to our further work.

REFERENCES

1. Paul E. Green. Jr, "Optical networking update", IEEE JSAC, 14, 764-779, 1996
2. David K. Hunter, et al, "WASPNET: A Wavelength Switched Packet Network", IEEE Comm. Mag., 120-129, March, 1999
3. M. Yoo, C. Qiao, "Optical Burst Switching (OBS): A new paradigm for an optical internet", JHSN, **8**, No. 1, 69-84, 1999
4. Ignacio de Miguel, Michael Duser, Polina Bayvel, "Traffic Load Bounds for Optical Burst-Switched Networks with Dynamic Wavelength Allocation", in ONDM 2001, vol. 1, Vienna, 2001
5. J. S. Turner, "Terabit burst switching", Journal of high-Speed Networks, Vol. 8, 1999, pp. 3-16
6. O. Gerstel., R. Ramaswami, G. Sasaki, "Cost-effective traffic groomingin WDM Rings", Proc. IEEE Infocom'98, pp. 69-77
7. S. Thiagarajan, A. K. Somani, "Performance Analysis of WDM Optical Networks with Grooming Capabilities", Proc. SPIE Boston, MA, USA, Nov. 2000
8. Kai Wu, Qingji Zeng, "Joint operation of packet network and WDM network", Proceedings of SPIE Vol. 4213, Boston, Nov. 2000