Optical Architecture for Multi-Terabit IP Routers

David K. Hunter, Ivan Andonovic
University of Strathclyde, EEE Department, 204 George Street, Glasgow G1 1XW, UK
Phone: +44 141 548 2527, Fax: +44 141 553 1955, Email: d.hunter@eee.strath.ac.uk

A novel architecture is proposed for future multi-terabit Internet Protocol routers, employing multiple cascaded stages of optical switching and buffering. External synchronization is not required, and a void-filling algorithm simplifies hardware requirements.

Optical Architecture for Multi-Terabit IP Routers

David K. Hunter, Ivan Andonovic
University of Strathclyde, EEE Department, 204 George Street, Glasgow G1 1XW, UK
Phone: +44 141 548 2527, Fax: +44 141 553 1955, Email: d.hunter@eee.strath.ac.uk

1. Introduction

Optical packet switching, to date exclusively concentrated on fixed-length packets, has been the subject of growing interest recently [1], with the intention of overcoming the anticipated problems inherent in constructing future very large electronic switch cores. Here these concepts have been extended to the design of optical packet switched node architectures suitable for use as IP (Internet protocol) routers, switching and buffering variable-length optical packets transparently. Each node may accept or transmit multiple optical packets simultaneously on each input or output fiber using wavelength division multiplexing (WDM), enhancing its throughput. The IP and WDM layers may be combined in future networks, simplifying network management and producing a highly flexible and functional packet switching layer. Simulation is used to study the performance of the proposed node under both bursty and self-similar traffic [2], the latter providing a meaningful comparison with real traffic. Although switching elements of any size can in principle be constructed, 16 inputs and outputs are the focus of the study.

2. Node Design Principles

High hardware complexity is a potential difficulty when implementing optical IP routers. Four measures are implemented to combat this:

- asynchronous operation, obviating the need to synchronize packets to timeslot or byte boundaries, prior to entering the switch thereby reducing the hardware overhead. It is assumed that packet lengths are multiples of one byte, while packet inter-arrival times are continuously distributed.
- statistical multiplexing among wavelengths is used to assist in contention resolution, reducing the buffering requirement [3]. If there is contention because two or more packets are directed to the same output simultaneously, initially an attempt is made to transmit each on a different wavelength. Buffers are still implemented in case the supply of wavelengths is exhausted.
- if the lowest increment of delay permitted is less than the minimum packet size, packets may be directed in a FIFO manner to the appropriate outputs; the hardware complexity inherent in this approach may however be avoided by using a technique, known as void filling [4].
- by use of multiple buffer stages in cascade, exponential increases in buffer depth for each additional stage can be achieved, an approach already proposed for fixed-length optical packets [5].

3. Node architecture

The architecture is composed of a series of S stages, numbered 0 to S-1 (Figure 1). Each stage has the following number of inputs and outputs: Stage 0: N inputs and D outputs; Stage S-1: D inputs and N outputs; All other stages: D inputs and D outputs.

Each pair of adjacent stages are interconnected by delay-lines which can subject each packet to a delay in $\{0, \delta N^{S-1-i}, 2\delta N^{S-1-i}, ..., (D-1)\,\delta N^{S-1-i}\}$, with δ being the smallest unit of delay, known as the delay-line granularity, i denotes the stage before the delays, with $i = 0$ representing the leftmost stage, and N is the number of inputs and outputs. Each link or delay-line within the architecture can carry multiple packets at once, by means of WDM.

Each packet entering a stage may leave on any other stage output, on any free wavelength. Hence each stage functions much like a crossconnect with wavelength conversion, only it operates at the packet rate. Λ is the maximum number of wavelengths per switching stage port, delay-line or link inside the

architecture. Throughout, the number of wavelengths on each node input or output (external to the architecture) is λ.

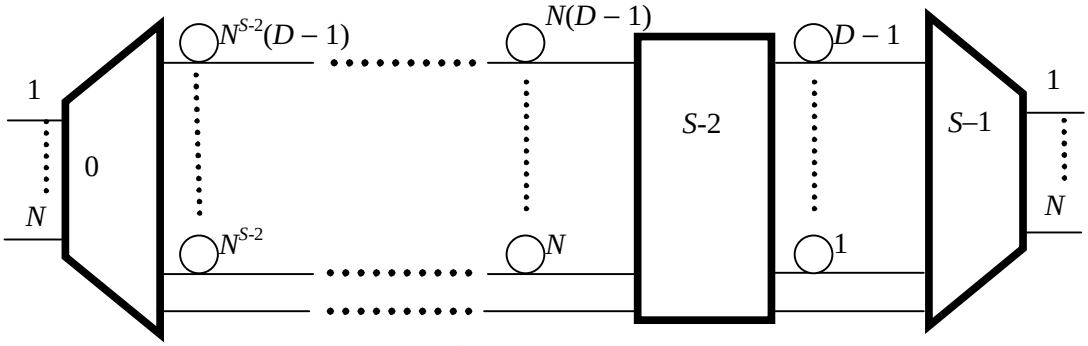

Fig. 1. Outline of the switch architecture with S stages.

4. Control Algorithm

The algorithm implements a modified form of the original void-filling algorithm [4], making a sequential search of the available routes. Packet priorities have not been examined, although this feature could be added. The algorithms are recursive; each time the algorithm is called to route a packet from a certain stage to the output, it calls itself to route the same packet from the next stage to the output, unless the stage is the final stage. The control algorithm maintains a list of the packets that are scheduled to pass through each point in the architecture in the future. The algorithms determine if a packet can take a particular path through the architecture, depending on whether there is contention with one or more packets that have been already scheduled, with the timing dictated by the delays in that path. If the path is free, then these lists are amended to reflect the allocation the packet to the route through the architecture. It is envisaged that the control unit will be implemented using VLSI technology.

5. Simulation Methodology

Two types of traffic are studied: traffic with a self-similar Pareto distribution and traffic with negative exponentially distributed inter-packet gaps and geometrically distributed packets. The packets were uniformly distributed across the outputs. Throughout, all packet lengths are measured in units of bytes. Simulations written in C, exhaustively tested against existing results e.g. [6] were carried out for 1.6×10^{8} packets to determine delay and packet loss above 10^{-6}, with a smaller packet loss being considered acceptable. The loading level was 0.8 per wavelength throughout.

For negative exponential traffic, even with just two stages, 32 wavelengths throughout, 16 delay-lines per stage, and 16 inputs and outputs, the packet loss was always less than 10^{-6}.

Simulations were carried out using self-similar traffic mimicking real traffic, corresponding to a Hurst parameter of 0.9. The minimum packet length was 400 bytes. Each simulation is carried out with the basic delay-line unit equal to 400, 2000, or 5000 bytes. Due to the deleterious effect of self-similar Pareto traffic, an appreciable packet loss was still obtained for two stages with such traffic.

6. Results: Self-Similar Traffic

Figures 2 shows the packet loss probability against number of delay-lines between stages, for 32 internal wavelengths and delay-line granularities from 400 to 5000 bytes. For two stages, it can be shown that the number of internal wavelengths has little influence on performance, while the larger values of delay-line granularity produce superior performance. This is because the total amount of delay-line storage available is greater, and void filling allows it to be utilized even although not all storage locations within it are immediately accessible.

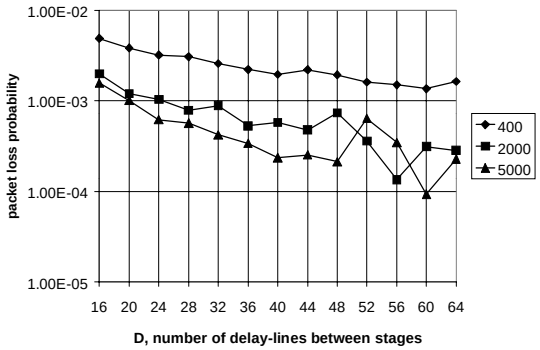

Fig. 2. Packet loss probability under self-similar traffic for two stages with 32 internal wavelengths. The numbers in the box indicate the delay-line granularity.

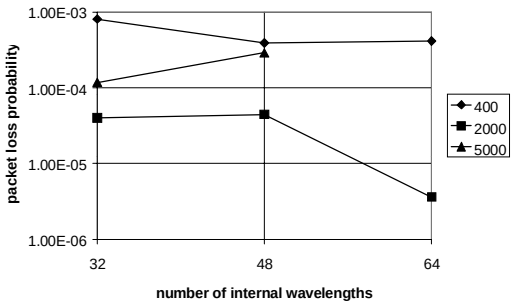

Fig. 3. Packet loss for three stages under self-similar traffic with 16 delay-lines between stages, and 32 external wavelengths.

Due to limits in simulation time, fewer simulations were carried out for three stages. Figure 3 shows the results for delay-line granularities of 400 to 5000, and 16 delay-lines between stages; the missing point in Fig. 3 indicates a packet loss probability of less than 10^{-6}. The results show that for 3 stages, 64 wavelengths and a delay-line granularity of 5000, a satisfactory packet loss of less than 10^{-6} was achieved.

Consideration of the delay for various configurations of the switch which yield a feasible packet loss of less than 10^{-6} shows increased delay penalties for a more economical architecture with more stages.

7. Conclusions

A new optical architecture has been presented for routing variable-length optical packets (e.g. IP datagrams), based upon asynchronous operation, the use of wavelength to resolve contention, void filling and a multi-stage architectural concept. The architecture was simulated under self-similar Pareto traffic with a Hurst parameter of 0.9. A number of conclusions can be drawn. For only 3 stages, when switching Pareto traffic, the architectures here can produce a useful packet loss of below 10^{-6}. Increasing the number of delay-lines between stages or increasing the number of stages improves the packet loss performance. Increasing the number of stages can reduce the overall number of delay lines and decrease the amount of hardware required to switch packets to the correct delay-lines. The performance of the architecture is improved by making the delay-line granularity much greater than the minimum packet length. Finally, the number of internal wavelengths has more influence on the performance of the 3-stage architecture, where there is more opportunity for internal blocking with a low number of internal wavelengths.

8. References

1. D. K. Hunter, M. C. Chia, I. Andonovic: "Buffering in Optical Packet Switches", *IEEE/OSA Journal of Lightwave Technology*, vol. 16, no. 12, December 1998, pp2081-2094
2. L. Tancevski, A. Bononi, L. A. Rusch: "Output Power and SNR Switching in Cascades of EDFAs for Circuit- and Packet-Switching Optical Networks", *IEEE/OSA Journal of Lightwave Technology*, vol. 17, no. 5, May 1999, pp733-742
3. D. K. Hunter, M. H. M. Nizam, K. M. Guild, J. D. Bainbridge, M. C. Chia, A. Tzanakaki, M. F. C. Stephens, R. V. Penty, M. J. O'Mahony, I. Andonovic, I. H. White: "WASPNET – a Wavelength Switched Packet Network", *IEEE Communications Magazine*, March 1999
4. L. Tancevski, A. Ge, G. Castanon, L. Tamil: "A New Scheduling Algorithm for Asynchronous, Variable Length IP Traffic Incorporating Void Filling", *OFC '99*, San Diego, CA, February 1999
5. D. K. Hunter, W. D. Cornwell, T. H. Gilfedder, A. Franzen, I. Andonovic: "SLOB: a Switch with Large Optical Buffers for Packet Switching", *IEEE/OSA Journal of Lightwave Technology*, vol. 16, no. 10, October 1998, pp1725-1736
6. L. Tancevski, L. Tamil, F. Callegati: "Non-Degenerate Buffers: An Approach for Building Large Optical Memories", *IEEE Photonics Technology Letters*, vol. 11, no. 8, August 1999