2013 18th OptoElectronics and Communications Conference held jointly with 2013 International Conference on Photonics in Switching (OECC/PS)

TuPT-4

An Optical Packet Switch Using Forward-Shift Switched Delay Lines

J. Touch*, S. Suryaputra*, J. Bannister**, A.E. Willner***

* USC/Information Sciences Institute, 4676 Admiralty Way, Marina del Rey, CA 90292, U.S.A.
** The Aerospace Corporation, 2310 E. El Segundo Blvd., CA 90245, U.S.A.
*** Ming Hsieh Department of Electrical Engineering, University of Southern California, 3740 McClintock Ave., Los Angeles, CA 90089, U.S.A.

Abstract

A 32x32 optical packet switch design using only four packets of variable delay is shown 95% as efficient as electronic switching using simulated Poisson Internet traffic. Our forward-shift approach is 10-30% better than a backward-shift.

I. Introduction

Optical routers are needed to support very high data rates (>10 Gbps) while avoiding costly conversion to electrical formats. All-optical packet routing raises significant challenges, including the need to process and restore digital optical signals, and for efficient packet multiplexing. Our design shows that only four packets of optical switched-delay line can multiplex packets nearly as efficiently as electronic random access storage.

Internet routers are packet switches that forward variable-length messages using four operations: forwarding lookup, hop-count decrement, checksum calculation, and packet multiplexing [1,2]. We are exploring the feasibility of all-optical implementations of each of these operations. We previously demonstrated optical forwarding lookup and hop-count decrement [3,4]; we are now developing optical checksums.

A major impediment to efficient packet multiplexing has been the perceived need for large amounts of random-access storage. Such storage is used as a “parking lot”, where packets from different “roads” are parked and separately retrieved. Current electronic approaches rely on such storage, and there is no optical equivalent.

Previous attempts to multiplex packets optically relied on input scheduling and sequential buffering [5,8,9] or recirculating buffering [10]. Our design uses switched-delay lines (SDLs) and approaches the efficiency of electronic switches [5]. This emulates highway merging, where one car speeds or slows to avoid collision.

We assume optical inputs and outputs each consist of a single channel of Internet packets. We forward packets independently, rather than as groups within circuits (as in Generalized Multiprotocol Label Switching) or trains of packets (as in Optical Burst Switching) [6,7].

II. Design

A packet switch consists of an input stage connected to an output stage by a switching fabric. The input stage filters traffic and selects its output stage; the output stage updates headers, schedules transmission, and adds link-layer headers (for multipoint links). The switching fabric delivers traffic between the input and output stages.

Packet switches are differentiated by the configuration of the stages and the switching fabric. A typical “demultiplexer/multiplexer” (demux/mux) design has N^2 internal links as a fabric, where the inputs demux traffic and the outputs mux to resolve contention (Fig. 1).

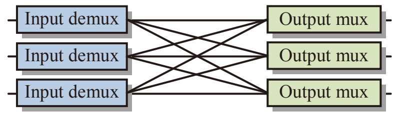

Fig. 1. Packet switch based on a demux/mux architecture.

Electronic switches use this architecture, in which virtual output queuing (VOQ) buffers packets in random-access memory at the input stage. The output stage schedules packets for transfer between the input stage queue and the output over a fixed-size cell switching matrix that emulates full internal connectivity. Two popular VOQ scheduling algorithms are Parallel Iterative Matching (PIM) and iterative 'slip' (iSLIP) [11].

All-optical switches cannot use VOQ approaches because there is no optical equivalent to random-access storage. Our design relies instead on optical FIFO queues implemented using SDLs, with delay of only four of the largest typical Internet packets (1500 bytes).

Our switch uses this demux/mux design in which each output stage is a lookahead variable-shift mux with separate optical data paths (thick blue) and electronic control paths (thin black) (Fig. 2) [12]. The mux uses two equal-length FIFO delay lines: the lookahead region provides fixed delay for batch scheduling; the shift region has configurable delay to resolve output contention.

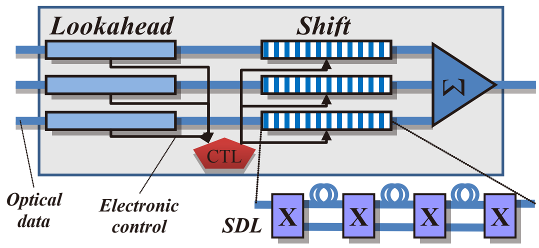

Fig. 2. SDL-based variable-shift mux.

Our approach differs from VOQ in two ways: our packets are queued at the output rather than the input, and our SDL FIFOs can speed up or slow down packets but cannot time-reorder them – and thus cannot support random-access, as would be needed for VOQ.

Traffic arrives from each input stage and enters the lookahead region, where an electronic controller (CTL)

This work is partially supported by the NSF Center for Integrated Access Networks (CIAN) contract 0812072.

schedules packets in batches, 48,000x slower than the optical packet data rate (every four 1500-byte packets). The lookahead region extracts header information used to configure the shift region. The SDL delays packets by default (taking the top, longer SDL path, Fig. 2); a forward shift occurs when the delay is shortened so that packets take the faster, direct path (through the bottom, shorter SDL path, Fig. 2). The same approach supports backward-shift, in which the fast-path is the default. Both SDLs can shift coarsely, *e.g.,* by 1/10 of a packet.

Packets are scheduled to optimize throughput and consider quality-of-service or other scheduling policies (Fig. 3, top). The schedule determines the SDL switch configuration, which delays or drops packets. The resulting batch has no contention, and the paths are passively merged to the output (Fig. 3, bottom).

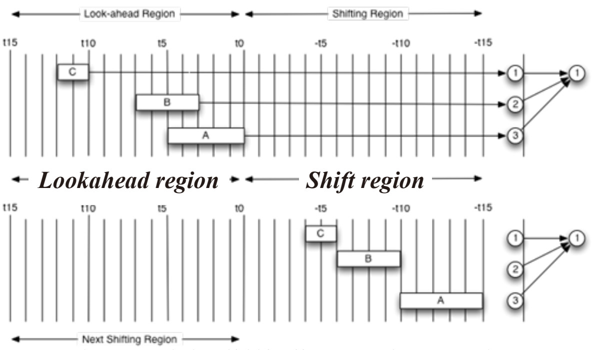

Fig. 3. Lookahead/shift effect on packet processing.

III. Simulation Results

We evaluated a 32x32 switch in using simulated Internet traffic generated from parameters derived from real traces. We compared an unbuffered switch to electronic VOQ-PIM, electronic VOQ-iSLIP, our optical forward-shift, and an optical backward-shift (inspired by [13]). Output throughput was determined under 100% input load using uniform random output port selection with Poisson (independent, typical for core routers) distributions of Internet packets using bimodal Internet packet sizes of 80% 40-byte and 20% 1500-byte packets (*i.e.*, in which roughly 10% of the bytes belong to small packets and 90% belong to large packets) [14,15], with SDL buffer lengths ranging from 1 to 100 packets.

The unbuffered configuration randomly drops packets that collide at outputs. Our earlier work showed that even an "omniscient" scheduler (selecting drops for maximum throughput by viewing the entire data stream) was only a few percentage points better than random [16]. For both forward-shift and backward-shift, our scheduler maximized throughput by shifting packets in order of arrival if they "fit" and dropping them otherwise.

Performance was measured using a custom C++ simulator. Each plotted point represents an average of runs with a standard deviation under 1% with 95% confidence (too small to plot). Each run represents 1 real second, and was dominated by steady-state behavior.

Optical forward-shift switching tracks the lower edges of electronic VOQ and achieves 10-30% higher throughput than backward-shift because packets use idle gaps more efficiently (Fig. 4). An unbuffered switch yields 50% throughput, as expected. Electronic VOQ perform best, but requires random-access storage and cannot be implemented in optics. VOQ has "stair-step" curves because they buffer in units of entire packets, so a larger queue is useful only when another entire packet fits. Variable-shift FIFOs have smooth curves because they shift by packet fractions.

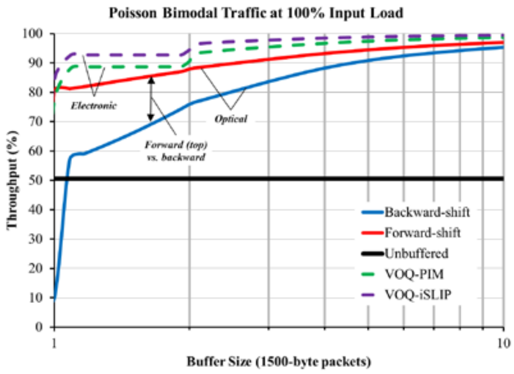

Fig. 4. Throughput comparison of 100% input load Poisson traffic.

IV. Conclusions

The need for large random access storage should no longer be an impediment to exploring all-optical packet switching. Forward-shift optical SDLs achieve 93.5% throughput, 95% as good as electronic VOQ, using only four packets of delay. Packets can be scheduled using an electronic control plane, in batches running 48,000x slower than line bitrate.

References

[1] F. Baker, (Ed.), RFC 1812, Jun 1995.
[2] S. Deering, R. Hinden, RFC 2460, Dec. 1998.
[3] M. Hauer, J. McGeehan, S. Kumar, J, Touch, J. Bannister, E. Lyons, C. Lin, A. Lau, H. Lee, D. Starodubov, A. Willner, *J. Lightwave Tech*., Dec. 2003, p.2765-2778.
[4] J. McGeehan, S. Kumar, D. Gurkan, S. M. Nezam, J. Bannister, J. Touch, A. Willner, *J. Lightwave Tech*., Dec. 2003, p.2746-2752.
[5] Z. Haas, *J. Lightwave Tech.*, May 1993, p.925-926.
[6] L. Berger, D. Fedyk, RFC 6002, Oct. 2010.
[7] C. Qiao, M. Yoo, *J. High Speed Networks*, Jan. 1999, p.69-84.
[8] S.Y. Liew, G. Hu, H.J. Chao, *J. Lightwave Technology*, Apr. 2005, p.1586.
[9] D.K. Hunter, W. Cornwell, T. Gilfedder, A. Franzen, I. Andonovic*, J. Lightwave Tech*., Oct. 1998, p.1725-1736.
[10] W. Xiaoliang, 22nd Int'l. Conf. on Advanced Information Networking and Applications, 2008, p.912-917.
[11] N. Mckeown, T. Anderson, *Computer Networks and ISDN Systems*, Dec. 1998, p.2309-2326.
[12] J. Touch, J. Bannister, S. Suryaputra, U.S. Patent #8,306,047, Nov. 2012.
[13] H. Harai, M. Murata, *Trans. on Networking*, V14 N1, Feb. 2006, p.191-204.
[14] R. Sinha, C. Papadopoulos, J. Heidemann, USC/ISI Tech. Report 643, May 2007.
[15] M. Crovella, M. Taqqu, A. Bestavros, In *A Practical Guide to Heavy Tails: Statistical Techniques & Applications*, (Birkhäuser, 1998) p.3-25.
[16] S. Suryaputra, J. Touch, J. Bannister, Proc. IEEE High-Speed Networks Workshop, Apr. 2009.