

2000

IP forwarding alternatives in cell switched optical networks

P. Boustead

University of Wollongong, boustead@uow.edu.au

J. F. Chicharo

University of Wollongong, chicharo@uow.edu.au

Publication Details

This paper originally appeared as: Boustead, P & Chicharo, J, IP forwarding alternatives in cell switched optical networks, IEEE International Conference on Communications, 18-22 June 2000, vol 3, 1628-1632. Copyright IEEE 2000.

IP forwarding alternatives in cell switched optical networks

Abstract

Optical switching will enable core Internet packet switching to scale with future transmission rate increases. Currently proposed optical ATM switches do not allow packet reassembly, which is necessary for packet level forwarding. This results in the requirement to create end to end ATM virtual connections for flows even if they contain only one packet. In electronically switched networks MPOA and MPLS allow both cell and packet level forwarding to overcome this problem. This paper examines the feasibility of implementing such protocols over an optically switched network. Two different architectures are examined: use of an adjunct electrical router; and native optical packet reassembly. An examination of the optical reassembly buffer requirements show that the use of MPLS will require significantly more buffering than MPOA.

Disciplines

Physical Sciences and Mathematics

Publication Details

This paper originally appeared as: Boustead, P & Chicharo, J, IP forwarding alternatives in cell switched optical networks, IEEE International Conference on Communications, 18-22 June 2000, vol 3, 1628-1632. Copyright IEEE 2000.

IP Forwarding Alternatives in Cell Switched Optical Networks

Paul Boustead, Joe Chicharo
Switched Networks Research Centre
University of Wollongong
Wollongong, Australia, 2522

Abstract—Optical switching will enable core Internet packet switching to scale with future transmission rate increases. Currently proposed optical ATM switches do not allow packet re-assembly, which is necessary for packet level forwarding. This results in the requirement to create end to end ATM virtual connections for flows even if they contain only one packet. In electronically switched networks MPOA and MPLS allow both cell and packet level forwarding to overcome this problem. This paper examines the feasibility of implementing such protocols over an optically switched network. Two different architectures are examined: use of an adjunct electrical router; and native optical packet re-assembly. An examination of the optical re-assembly buffer requirements show that the use of MPLS will require significantly more buffering than MPOA.

I. INTRODUCTION

Optical transmission capacity is increasing dramatically with the introduction of Wave Division Multiplexing (WDM). In order for future switching architectures to keep up with projected transmission capacities there has been much interest in the development of optical switching technology. The synchronous time-slotted nature of proposed optical switching architectures means that an ATM data-link layer is likely. It is therefore important to consider packet over ATM forwarding alternatives.

The current protocols for IP over cell in the electrical domain, such as Multi-Protocol Over ATM (MPOA) [1] and Multi-Protocol Label Switching (MPLS) [2], allow some network layer and some cell level forwarding in the network. Allowing network layer forwarding within core switches eliminates the need to create end to end ATM connections for short packet flows. The important issue of quantitatively examining the implementation of such protocols over optical switches does not appear to have been addressed in literature thus far. Protocols such as MPOA and MPLS were designed and optimized for use with traditional electrical switching/routing technology. The nature of optical switching hardware necessitates different design criteria for an IP over ATM protocol. It is particularly important to reduce complexity, and to minimize the amount of optical buffering. A reduction of output buffer requirements may be possible using traffic smoothing techniques, for example, [3] proposes a mechanism that allows lossless optical switching with small output buffers. However, these techniques will not reduce the re-assembly buffer required by aggregated label switching techniques that use VC-Merge. This paper investigates the possible implementation of these proposals in an optically switched network.

We examine the forwarding mechanisms of several IP over cell approaches including MPOA, MPLS and IP Switching. Two methods of supporting IP over optical cell switches are examined: Use of a simple optical cell switch with an ad-

adjunct electrical router, and use of an optical switch that supports packet reassembly in the switch fabric. Use of an adjunct router will minimise switch buffering but will not support MPLS. The second case is examined in detail with a trace driven simulation. Of particular interest is the amount of buffering required per port for packet reassembly. Packet reassembly is required in varying degrees by each approach for network layer forwarding as well as ensuring cell sequencing of ATM Adaptation Layer 5 (AAL5) data streams. We perform a discrete-event simulation analysis to compare the size of reassembly buffers required by the different protocols. We find that the use of aggregated packet forwarding protocols such as MPLS requires on average twice the number of cells for reassembly buffers than nonaggregated protocols such as IP Switching and MPOA.

The next section introduces aggregated and nonaggregated forwarding mechanisms. Section III discusses the issues related to packet forwarding using optical cell switches. The simulation used to compare aggregated and nonaggregated forwarding is discussed in Section IV. Section V presents optical re-assembly results. Adjunct router results are presented in Section VI. Section VII concludes the paper.

II. PACKET FORWARDING TECHNIQUES

Current packet over cell forwarding techniques are designed to improve the scalability of electronic switch routers. IP forwarding is bypassed, for a large percentage of packets, by dynamically created cell switched paths. There are several different protocols that have been developed to do this including MPOA, MPLS, and IP Switching. These protocols have substantially different mechanisms for creating cell switched paths. This section examines these mechanisms. Of major interest is the amount of buffering required. The type of packet forwarding mechanism will have little or no effect on the output buffer required (work aiming to reduce output buffer sizes can be seen in [3]). However, packet forwarding mechanisms will have a significant effect on the size of reassembly buffers required.

Re-assembly buffers are required by packet forwarding mechanisms for two purposes: IP forwarding, and to ensure ATM Adaptation Layer 5 (AAL5) cell sequence integrity. We group packet forwarding protocols into two groups, nonaggregated and aggregated, depending upon the need for cell stream merging. Non-aggregated approaches include IP Switching and MPOA, and require reassembly buffers only for IP forwarding. Aggregated approaches such as MPLS and Tag Switching require reassembly buffers for IP forwarding and to maintain AAL5 cell sequence integrity.

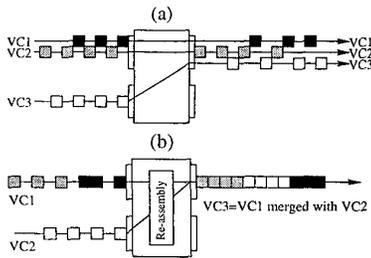


Fig. 1. Non-aggregated (a) and aggregated (b) packet forwarding

A. Non-Aggregated Forwarding

Non aggregated forwarding techniques maintain separate virtual circuits (VCs) for source/destination pairs as shown in Figure 1 (a). Packet reassembly is only required for the packets that are forwarded at the network layer. Examples include IP Switching, and MPOA.

IP Switching [4] and MPOA [1] determine if a cut-through flow should be created based on the level of traffic flow. The main difference between them, in the context of this comparison, is where the decision to cut-through is made. The ingress node of the MPOA network decides if an end-to-end cut-through is necessary. On the other hand, every switch in an IP switching network is involved in creating its local segment of the cut-through route. If the number of packets in a flow exceed a "packet threshold" in a certain time period (usually 60 seconds [4]) then the cut-through is created for that flow. These approaches make similar use of reassembly buffers. Re-assembly buffers are used solely for packet level forwarding, since cut-through flows are defined depending on source and destination addresses and VC-merge is not required. However, MPLS and other aggregated forwarding techniques approaches require additional reassembly buffers in order to maintain sequence integrity of AAL5 cells in the aggregated streams.

B. Aggregated Forwarding

Aggregated forwarding techniques merge one or more VCs from different input ports to a single VC on the output ports as shown in Figure 1 (b). The merging of VCs leads to the necessity for packet reassembly at VC merge points. This is due to the use of AAL5, which uses only an end of packet bit in the last cell of a segmented packet for delineation. If cells belonging to AAL5 encoded packets are interleaved then the packets can no longer be reassembled. Tag Switching [5] and Multiprotocol Label Switching (MPLS) [2] are both examples of protocols that support aggregated label-switching. Both approaches are similar and will be described together.

The trigger for the creation of a Tag Switching cell level cut-through is either the receipt of a standard IP routing protocol packet advertising a new route, or a proprietary tag (Tag Switching term for a label) distribution protocol packet [6]. The main component of a tag-switching network is a Tag Switch. A Tag Switch maintains a Forwarding Information Base (FIB), and a Tag Information Base (TIB). The FIB is populated using information from routing protocol messages, and

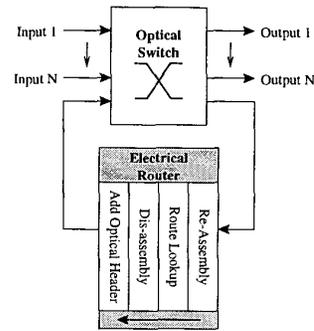


Fig. 2. Adjunct Router

is similar to the routing table in a standard IP router. The TIB is essentially the switch VC table. Tag Switches bind all entries in the FIB with tags in the TIB. The first hop Tag Switch performs network layer forwarding to find the correct entry in the FIB. The associated tag in the TIB will then be placed in the cell's VPI field, and the datagram is then forwarded through the ATM switch using this tag. Subsequent Tag Switches will have previously set-up bindings between this tag and a tag for its next hop router to the destination. Thus, ATM will switch the datagram to its destination.

III. OPTICAL PACKET FORWARDING

We examine two types of packet forwarding solutions in cell switched optical networks. The first method uses an adjunct electrical router to perform reassembly and higher level forwarding when required. The second method performs reassembly in the optical switch fabric to facilitate flow merging for aggregated and nonaggregated forwarding.

A. Adjunct Router

The first approach can be seen in Figure 2. Packets that require reassembly are switched through the optical switch to an adjunct electrical router. Within this router packets are converted to the electrical domain, reassembled, an IP forwarding decision is performed, dis-assembled into cells, and finally an optical header to route the cell to the correct output port is added. The cells are then rerouted through the switching fabric to the correct output port. AAL5 cell sequence integrity is maintained by sequential passage of routed packets through the electrical router to one input port of the cell switch.

This approach is applicable to the nonaggregated protocols since only a small proportion of packets require reassembly. However, for aggregated protocols reassembly in the form of VC merge (not IP forwarding) is required for all packets within merging streams. To implement aggregated protocols it must be possible to perform reassembly (VC Merge) within the optical switch which is the second option we consider. The main design requirement, of the adjunct router approach, is to minimise the use of the electrical router and switch a high percentage of cells optically. Results for the percentage of packets switched in the nonaggregated IP Switching protocol are presented in [4]. This indicates that the utilisation of the adjunct

router will be between 10% and 20% with a packet threshold of 10.

B. Optical Reassembly

The second option we consider is performing reassembly within the optical switch fabric. This approach would enable VC merging and therefore use of aggregated protocols such as MPLS. The reassembly ability would also allow nonaggregated approaches without using an adjunct electrical router. A feed-back optical buffer must be used since a feed-forward buffer can only be used if the time the cell is in the switch fabric is known when the cell enters the switch. We envisage using electronic control of optical fiber loop buffers to simulate reassembly buffers. This would be performed in a similar way to which the output buffer is simulated using the central fiber loop buffer and electronic control. This electronic control to "simulate" an output buffer is described in more detail in Section III-C.

Aggregated label switching protocols such as MPLS will require the label encoded in the optical header, and enough switch fabric buffering to reassemble all packets within merging streams. The nonaggregated approach will require labels or VC identifiers encoded in cells that belong to cut-through streams. Destination information is encoded in the optical header of the small percentage of cells not belonging to a cut-through path. We assume a fast IP lookup [7] for this small percentage of cells while they are being buffered.

If reassembly is performed in the optical domain then the most important metrics to measure are related to buffer usage. It is important to reduce the size of buffers, and to reduce the time that cells spend in buffers.

C. Optical Buffering Technology

There are many proposed optical buffering designs which can be divided into two broad categories: feed-forward and feedback buffered switches [8]. Cells entering a feed-forward buffer pass through a fixed number of optical delay lines. Feedback buffers have the capability of feeding cells back through delay lines multiple times. Of the two types of buffers the feedback buffer is the only one that is capable of packet forwarding with reassembly, using electronic control. The feedback buffer is able to hold a cell until the rest of the packet has been received, whereas the feed-forward switch must select a fixed delay as each cell is received.

An example of an optical switch that uses the feedback optical buffering concept is the fiber loop switch [8] in Figure 3. The switch buffering consists of a single cell period loop of fiber. Utilizing WDM the capacity of the buffer is m , where m is the maximum wavelengths available. When a cell enters the switch the header is converted to the electrical domain and used by the electronic control circuit that co-ordinates the optical switching and buffering. The optical data component of the cell is converted to a spare wavelength or "memory location" and enters the fiber loop. The electronic control maintains the cells in the loop for a time equal to a traditional output buffered switch. The optical data component is then switched to the appropriate output switch. There are other proposed feedback

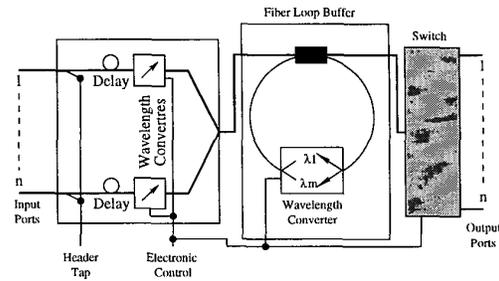


Fig. 3. Fiber loop switch [10]

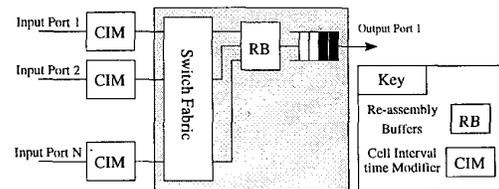


Fig. 4. Simulation block diagram

buffered switches such as the Shared Memory Optical Packet switch (SMOP) described in [9]. This approach uses delay loops of different lengths.

IV. SIMULATION DESCRIPTION

Discrete event simulation techniques were used to compare aggregated and nonaggregated packet forwarding. The aim of the simulation comparison is to investigate the average buffer usage for a given packet loss probability.

A block diagram of the simulation is shown Figure 4. A single core label switch is modeled with N input ports and a single output port. An output buffered switch was modeled since most optical switching designs simulate an output buffered switch [8]. Traffic arriving at each input port enters the Cell Interval Modifier (CIM) block that varies the cell inter-arrival time. Re-assembly buffers and output buffers are located on the output side of the switch fabric. Cells that do not require reassembly bypass the reassembly buffers and are placed directly in the output buffer.

The simulation is fed by a packet level traffic trace obtained from the National Laboratory for Applied Networks Research (NLNR) trace number 960228. The traffic trace consists of 10.6×10^6 packets over a period of 770 seconds. This choice of traffic trace enables us to validate nonaggregated simulation outputs for percentage of packets switched and VC usage with [4]. The trace is used to represent aggregated traffic on the output port of the simulated switch. Traffic for individual input ports is obtained by dividing the traffic equally between ports using the IP source address for each packet flow. The packets are divided into ATM cells with the cell inter-arrival time determined by the CIM block. The CIM block inserts an average cell inter-arrival time for cells representing each IP packet. The average cell inter-arrival time is an input to the simulation. We chose an average cell interval of 10 cells for most tests. This

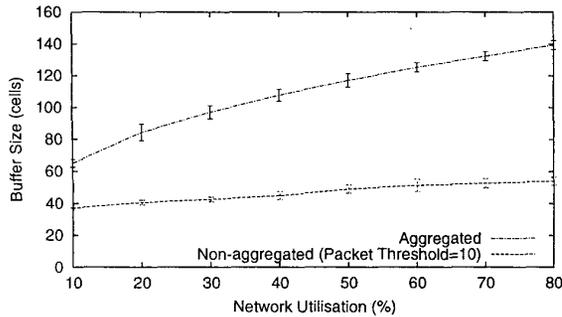


Fig. 5. Average buffer size versus utilisation for 10^{-4} packet loss probability

cell inter-arrival time is used to simulate the switch servicing other VC's and other priority traffic. This parameter is varied to examine its effect on the reassembly buffer size required.

Non aggregated approaches are modeled using separate VCs for each source and destination pair. The reassembly buffer was therefore only required for network layer forwarding. We ignore connection setup delays in order to concentrate on the effect of varying packet threshold.

We are interested in the number of cells required for reassembly buffers to forward cells with low packet loss probability. In order to do this we measure the number of cells in the reassembly buffer as each packet is forwarded. The simulation provides an average as well as cumulative distribution of reassembly buffer size. The cumulative buffer size distribution is used to determine buffer size requirements for a given packet loss. The average packet size is also examined to determine if there are differences between sizes of switched and routed packets. The size of routed packets will affect the reassembly buffer sizes that are required.

V. OPTICAL RE-ASSEMBLY RESULTS

This section examines the optical reassembly buffer requirements of various packet forwarding techniques. The major finding is that aggregated approaches require approximately twice the buffering of nonaggregated protocols for the same packet loss probability. This result is due to a significant reduction in the number of packets that require reassembly, as well as a significant difference in the size of routed and switched packets for the nonaggregated approach.

This section will first provide a comparison of buffer sizes for aggregated and nonaggregated approaches. This is followed by an examination of the performance of nonaggregated approaches concentrating on the selection of the packet threshold parameter. The effect of varying the cell gap is also examined.

A. Re-assembly Buffer Size

The required buffer size for a packet loss probability of 10^{-4} (bound by trace length) is shown in Figure 5. Even at low network utilization the aggregated approach requires significantly more buffering. At a network utilization of 50% aggregated packet forwarding requires an additional 125% buffering over the nonaggregated approach.

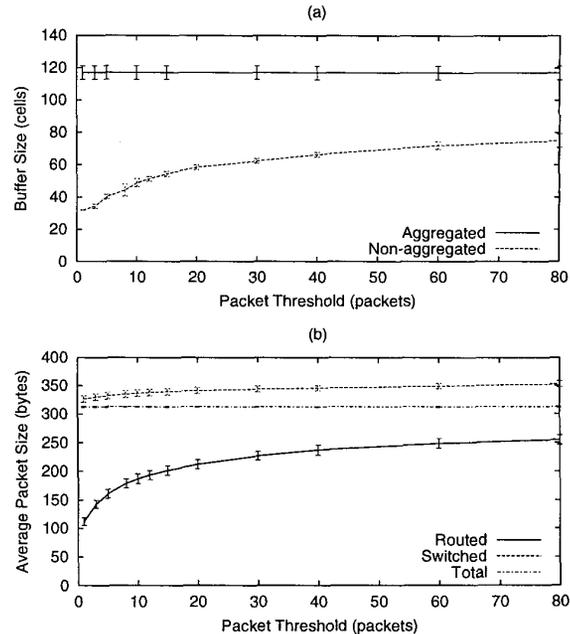


Fig. 6. Effect of packet threshold on (a) re-assembly buffers (b) packet sizes

B. Performance of nonaggregated packet forwarding

This section examines the reasons for the superior performance of nonaggregated packet forwarding protocols. The main parameter controlling the performance is the packet threshold. Varying this parameter controls the percentage of packets forwarded at the network layer and will have an effect on the reassembly buffer requirements.

A packet threshold of 10 was chosen by [4] as a sensible tradeoff between buffer usage and VC usage. However, we are more interested in minimizing optical buffer usage than electrical memory so we examine the tradeoff between packet threshold and optical buffer usage. In Figure 6 (a) we show the relationship between packet threshold and buffer usage for a packet loss probability of 10^{-4} . The buffer usage for the aggregated approach of 118 cells is shown for comparative purposes. At a packet threshold of one an average buffer size of 30 is required, this increases to 75 at a packet threshold of 80. The "knee" of the curve is at a packet threshold of 10.

The average packet sizes for switched and routed packets versus the packet threshold, before a cut-through is created, can be seen in Figure 6 (b). The choice of packet threshold has a significant effect on average packet sizes of routed packets. With a packet threshold of 5 the average size of routed packets is 160 bytes (in this trace) this is 57% less than the average size of switched packets. At a packet threshold of 80 the average size of routed packets increases to 250 bytes. A reduction in the size of routed packets will result in smaller reassembly buffer requirements for nonaggregated forwarding, as seen in the buffer size results in Section V-A. It is interesting to note the similarity between the average buffer size curve for nonag-

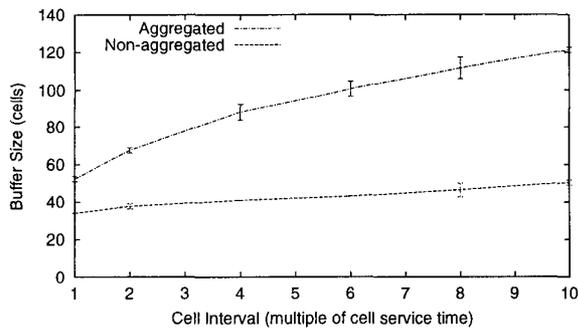


Fig. 7. Buffer size versus cell interval

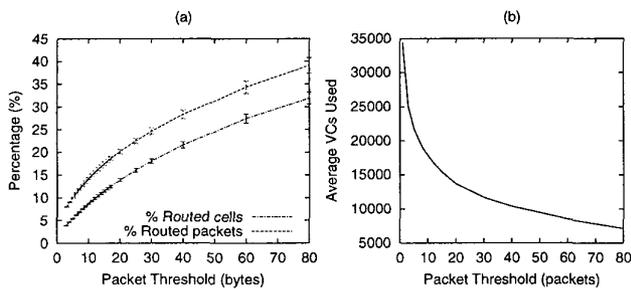


Fig. 8. (a) Percentage cells switched and (b) VC usage versus Packet threshold

gregated approaches in Figure 6 (a) and the average packet size for routed packets in Figure 6 (b).

The effect of varying the gap between cells, with a switch utilisation of 50%, is shown in Figure 7. Even with a low cell gap the nonaggregated approach requires significantly smaller buffers (50% less than the aggregated approach). Higher cell gaps may be introduced by scheduling mechanisms handling different levels of priority traffic. At higher cell gaps the difference in performance of aggregated and nonaggregated forwarding increases. With a cell gap of 10 cell service times the nonaggregated approach requires a buffer of 50 cells while the aggregated approach requires 120 cells, which represents an increase of 140%. Clearly, aggregated forwarding is significantly more sensitive to an increase in cell gap.

VI. ADJUNCT ROUTER RESULTS

Use of the adjunct router approach does not require additional cell buffering in the switch core. However it requires processing of a proportion of cells in the adjunct electrical router. It is important to reduce the percentage of cells processed by this router. The NLNR trace used in the optical buffering simulation was also used to determine the percentage of cells/packets switched for the nonaggregated IP Switching protocol as done in [4].

The percentage of packets and cells switched with a varying packet threshold are shown in Figure 8 (a). The packet threshold value also has a significant effect on the number of VCs which is shown in Figure 8 (b). The packet threshold value 10 is commonly chosen [4] as a compromise between VC usage

and percentage of packets switched. At a packet threshold of 10 the percentage of packets forwarded by the adjunct router is 14% which corresponds to 8% of packets switched. The percentage of cells routed is lower than the percentage of packets routed because, on average, the size of routed packets is significantly smaller than the overall average packet size (as shown in Figure 6 (b)). This can be reduced to 5% of packets switched if a packet threshold of 5 is chosen.

VII. CONCLUSIONS

This paper examines the feasibility of using current packet over cell protocols in an optically switched environment. Protocols are classified into aggregated and nonaggregated. Two different architectures were examined: use of an adjunct electrical router; and reassembly support within the optical cell switch.

Using an adjunct electrical router is only feasible for nonaggregated protocols such as IP Switching. The use of aggregated protocols precludes the use of the adjunct electrical router architecture.

The use of aggregated and nonaggregated protocols with an optical switch that supports reassembly was compared, by simulation, to establish reassembly buffer size requirements. We assume that output buffer requirements have been minimised using traffic smoothing mechanisms. A simulation comparison showed that aggregated forwarding requires significantly larger optical buffers than nonaggregated forwarding. The large difference was shown to be a result of a significant reduction in the average size and number of reassembled packets in the nonaggregated case. This reduction in packet size was found to be an artifact of the packet threshold mechanism.

These results indicate that nonaggregated protocols are significantly more suited to an optically switched environment.

REFERENCES

- [1] Andre Fredette. *Multi-Protocol Over ATM Version 1.0*. The ATM Forum, 1997.
- [2] R. Callon, P. Doolan, N. Fieldman, A. Fredette, G. Swallow, and A. Viswanathan. A framework for multiprotocol label switching. Internet draft, IETF, November 21 1997. Expires May 21 1998.
- [3] L.E. Moser and P.M. Melliar-Smith. Lossless packet switching with small buffers. *IEEE Proc. Commun.*, 143:335–340, 1996.
- [4] Peter Newman, Greg Minshall, and Thomas Lyon. Ip switching - atm under ip. *IEEE/ACM Transactions on Networking*, 6(2):117–129, 1998.
- [5] Yakov Rekhter, Bruce Davie, Eric Rosen, George Swallow, Dino Farnacci, and Dave Katz. Tag switching architecture overview. *Proceedings of the IEEE*, 85(12):1973–1983, 1997.
- [6] P Doolan, B Davie, D Katz, Y Reckhter, and E Rosen. Tag distribution protocol. Internet Draft draft-doolan-tdp-spec-01.txt, IETF Network Working Group, May 1997.
- [7] Mikael Degermark, Andrej Brodnik, Svante Carlsson, and Stephen Pink. Small forwarding tables for fast routing lookups. In *ACM Sigcomm'97*, 1997.
- [8] David K. Hunter, C. Chia, and Ivan Andonovic. Buffering in optical packet switches. *Journal of Lightwave Technology*, 16(12):2081–2094, 1998.
- [9] M. J. Karol. Shared-memory optical packet (atm) switch. *Proceedings SPIE: Multigigabit fiber communications systems*, 2024:212–222, 1993.
- [10] F. Masetti, J. Benoit, F. Brillouet, J. Gabriagues, A. Jourdan, M. Renaud, D. Bottle, E. Eilenberger, K. Wunstel, M. Schilling, and D. Chiaroni. High speed, high capacity atm optical switches for future telecommunication transport networks. *IEEE Journal on selected areas in communications*, 14(5):979–998, 1996.