Tbit/sec Switching Scheme for ATM/WDM High-Speed Computer Networks*

Itamar Elhanany[1], Dan Sadot[2]
Electrical and Computers Engineering Department
Ben-Gurion University
Beer-Sheva 84105, Israel

Abstract

A novel high throughput, reservation-based switch architecture for ATM/WDM networks is presented. The scheme is contention-free and highly flexible yielding a powerful solution for high-speed broadband packet-switched networks. Switching management and control is studied for data rates of up to 10 Gbit/sec/port, providing aggregated throughput of over 1 Terabit/sec.

1. Introduction

Wavelength division multiplexing (WDM) is widely recognized as a promising technology for high-capacity, scalable, cost effective optical networks. Considerable attention has been paid to WDM contention resolution and wavelength assignment issues, commonly targeted at packet-switched computer networks such as asynchronous transfer mode (ATM) (Chiussi *et al.*, 1997; McKinnon *et al.*, 1998). Recently, several switching schemes were designed in order to support high capacity, large number of ports and low latency requirements (Chiussi *et al.* 1997, Salisbury *et al.*, 1998; Sivaraman *et al.,* 1997). Here, we introduce a new WDM packet-switching scheme that offers Terabit/sec capacity along with support of over 100 ports and switching latency of several microseconds. The new architecture is contention-free and has the advantages of low implementation complexity and high scalability. In addition, the architecture can easily be adapted to comply with diverse quality of service (QoS) demands.

2. Switching Architecture

Figure 1 depicts the proposed switch architecture. The nodes, corresponding to the switch ports, have bi-directional optical data links interconnected via an optical passive star coupler. Each transmitter can be tuned to any of the N wavelengths, while each receiver is

* Work supported by the Israeli Ministry of Science under contract no. 9404-1-97, the Wolfson foundation and the Rector's office at the Ben-Gurion University of the Negev.

[1] E-mail: itamar@ieee.org
[2] E-mail: sadot@ee.bgu.ac.il

assigned a fixed and distinct wavelength. ATM traffic received at each port is distributed to different buffers within the node on a cell-by-cell basis, where each buffer relates to a single wavelength according to the desired destination node. All nodes are connected to a central wavelength reservation scheduler via a common electronic wavelength reservation bus, and individual control lines. The *N* bus lines are accessible to all nodes and indicate the reservation status of each of the *N* wavelengths. The individual control lines are used by the scheduler to signal each node, in turn, to commence the wavelength reservation procedure.

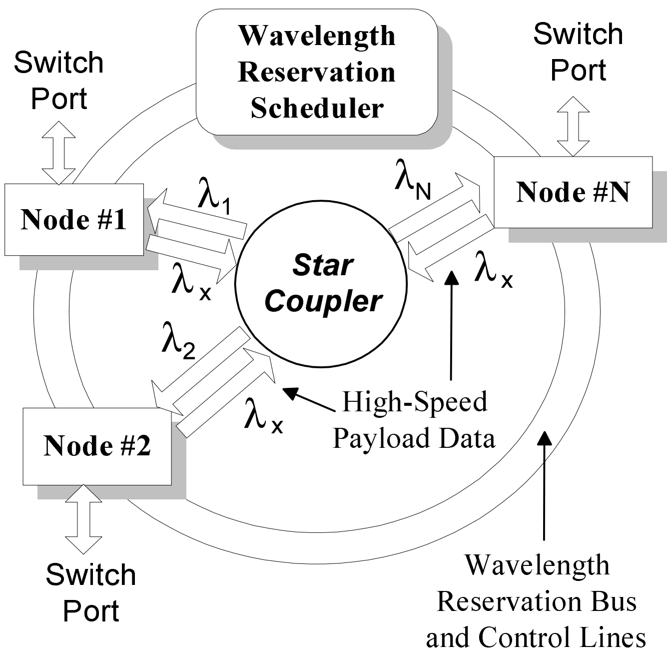

Figure 1: Terabit/sec ATM over WDM switch architecture

3. Wavelength Reservation

Upon receiving a control signal from the wavelength reservation scheduler, each node performs wavelength reservation according to two major guidelines: (a) global switch resources status, i.e., available wavelengths at the reservation time, and (b) local considerations, i.e., the status and priorities of the node's internal buffers.

Figure 2 illustrates a block diagram of the wavelength reservation hardware of a node in an 8x8 switch. At the highest level, the wavelength reservation status lines, denoted by S_i, either grant or discard buffer priorities, denoted by P_i, via designated AND logic. Consequently, only buffer indices, denoted by I_i, relating to available wavelengths advance to the lower levels. Each level consists of a set of comparators, which receive as input a pair of indices along with their respective priorities and output the higher priority and its corresponding index. The output of the last comparator determines the “prevailing” buffer, which held the highest priority out of the subset of buffers relating to unreserved wavelengths. Any number of parameters, such as buffer load, accumulated delay and required QoS can affect the buffer priority metrics. Node wavelength selection is instantly followed by assertion of the relevant line within the wavelength reservation bus. At that point, utilizing a weighted Round Rubin procedure, the central scheduler signals the next node to begin wavelength reservation. Contention is avoided, since at any given time only one node attempts to reserve a wavelength. After all *N* nodes complete wavelength reservation, high-speed data is optically transmitted via the star coupler.

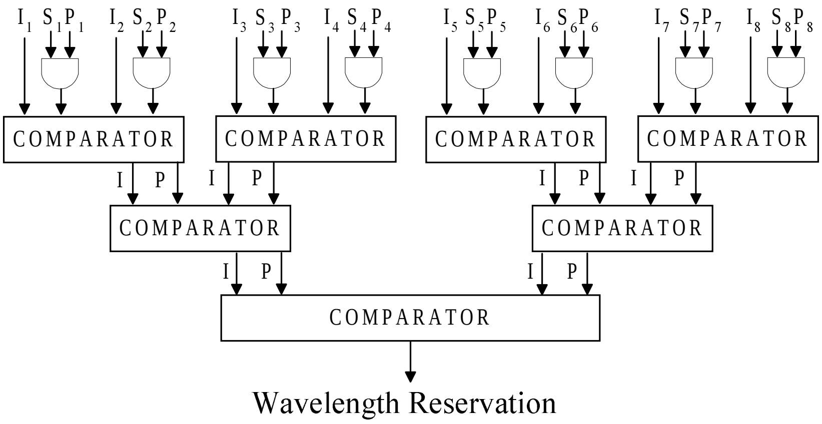

Figure 2: Block diagram of node wavelength reservation for an 8x8 switch.

The wavelength reservation and data transmission are conducted in a time slot discipline. Nodes transmit data concurrently at wavelengths reserved during the previous time slot. Assuming *N* nodes, the time slot period can be calculated as $t_{ts} = N \cdot \log_2(N) \cdot t_c$, where t_c is the processing time for a single comparator. Accordingly, t_{ts} dictates the minimal number of cells required to be transmitted during each time slot and hence the queuing time delays.

4. Simulation Results

Taking into consideration present custom VLSI technology, t_c=1 *ns* was found feasible. As a result, extremely short processing time for resource allocation is attained, yielding high switching performance. Figure 3 presents the number of cells required to be transmitted during each time slot under heavy traffic conditions, versus switch aggregated throughput. Various port bitrates are presented as a parameter. The marks on each curve represent standard number of ports, e.g., 8, 16, 32, 64 and 128. Simulations demonstrate that Terabit/sec switching capacity at 9.95 Gbps (OC-192) per port can be achieved. Over to 100 ports are supported with under 20 cells transmitted per time slot, and with total required buffer size of a few hundreds of cells.

These results can be compared to other high-speed buffer management approaches such as in (Chiussi *et al.*, 1997). In such architectures, buffer selection is executed using reservation-free Round-Robin algorithms that attempt to minimize processing time. Contention is avoided by applying multilayered buffer configurations and backpressure mechanisms, which limit performance and scalability.

Although in this letter we discuss fixed-size packets, this method can be applied to variable-size packets, e.g., IP traffic. By the same token, optical WDM can be replaced by other multiplexing technology, e.g., optical SDM (space division multiplexing).

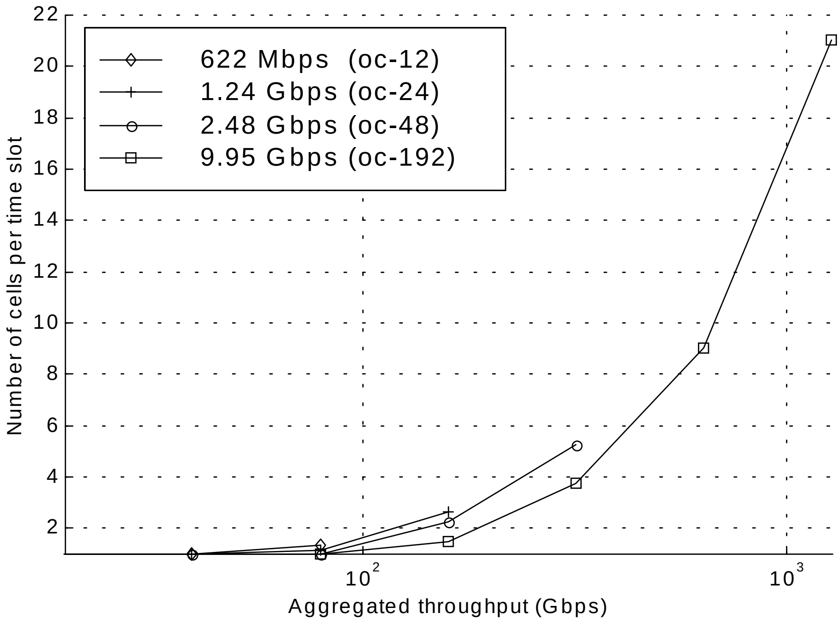

Figure 3: Minimal number of cells transmitted per time-slot for various aggregated throughput values

5. Conclusions

A novel Tbit/sec switch architecture for ATM over WDM packet-switched networks has been proposed. By applying an ultra-high speed scheduling discipline, extremely high capacity and low latency switching is achieved for fabrics of up to 128x128. The design is simple, scalable and flexible to support diverse traffic characteristics. The method can be adapted to IP traffic and to various multiplexing technologies.

References

Chiussi, F. M., J. G. Kneuer, and V. P. Kumar (1997). “Low-Cost Scalable Switching Solutions for Broadband Networking: The ATLANTA Architecture and Chipset,” *IEEE Communications Magazine*, **25**, no. 12, pp. 44-53.

McKinnon, M. W., N. R. George, and G. P. Harry (1998). “Performance Analysis of Single Hop Photonic ATM Switch Architecture with Tunable Transmitters and Fixed Frequency Receivers,” *Performance Evaluation*, **33**, no. 2, pp.113-136.

Salisbury, C., and R. Melhem (1998). “A High Speed Scheduler/Controller for Unbuffered Banyan Networks,” *Proc. IEEE ICC’98*, Session S18-6.

Sivaraman, V., and G. N. Rouskas (1997). “HiPeR-1: A High Performance Reservation Protocol with Look-Ahead for Broadcast WDM Networks,” *Proc. IEEE INFOCOM ’97*, **3**, pp. 1270-7.