Fairness Issues in an Embedded Photonic Ring Interconnect

Abhijit Mahajan
Mark Franklin
Roger Chamberlain

Abhijit Mahajan, Mark A. Franklin, and Roger D. Chamberlain, “Fairness Issues in an Embedded Photonic Ring Interconnect,” in *Proc. of 4th High Performance Embedded Computing Workshop*, September 2000.

Computer and Communications Research Center
Washington University
Campus Box 1115
One Brookings Dr.
St. Louis, MO 63130-4899

Fairness Issues in an Embedded Photonic Ring Interconnect*

Abhijit Mahajan Mark Franklin Roger Chamberlain
Computer and Communications Research Center
Washington University, St. Louis, Missouri

1 Introduction

Recent advances in VLSI photonic technologies enable us to design and implement optical interconnection networks with terabit per second bandwidth capacity. A physical ring architecture that exploits the high bandwidth provided by these new technologies is described in [1]. The design is targeted at multiprocessor systems that require frequent, massive data transfer among processors, from input sensors, and to output devices. Examples of embedded systems with these requirements include space-time adaptive processing, synthetic aperture radar, and real-time image processing applications.

At the heart of the architecture is a VLSI photonic device based on the use of an $M \times M$ array of Vertical Cavity Surface Emitting Laser (VCSEL) and detector pairs [3]. Each VCSEL-detector pair will be capable of operating at rates exceeding 1 Gbps. With $M = 32$, the raw bandwidth deliverable will be greater than 1 Tbps. This technology is currently only available as a custom design, however, commercial development is being pursued.

2 Multiring Architecture

The multiprocessor interconnect described in [1] utilizes a ring topology. Consider a four node example where each processing node is connect to the ring as shown in Figure 1. Given the numerous VCSEL-detector pairs, we can assign disjoint subsets of VCSEL-detector pairs to each processing node. If these subsets are allocated according to receiver designation, then each subset can be thought of as a channel associated with messages being received by a given node. This arrangement implements a multiring topology [2].

With the multiring topology, each channel can be thought of as a daisy chain terminating at the receiver. Figure 2 shows a four node (P1, P2, P3, and P4), four channel system. In the example, if node 4 wants to send a message to node 2, it will send the message on channel 2. The message will first be received on channel 2 of node 1's detector array. Node 1 will then repeat that message on channel 2 of its VCSEL array. The message is then reflected to channel 2 of node 2's detector array and is thus received by node 2.

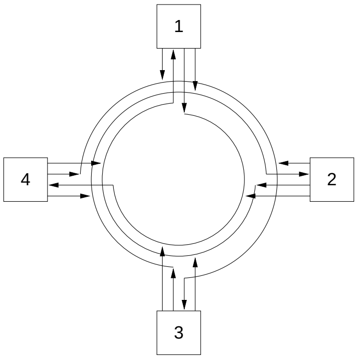

Figure 1: Multiring topology[2].

Since, in a multiring, node i will never send on channel i, there will always be an extra channel (marked extra in Figure 2). This extra capacity can be used for control channels (as opposed to the data channels). In [1], the control channels are used to implement reliable message delivery. Here, we describe their use in implementing a channel arbitration procedure that ensures fairness.

In the basic ring implementation, to minimize buffering requirements, whenever there is contention for a channel, the upstream node is granted access. This has the side effect of imposing a priority structure on the nodes in the system, which can result in non-uniform access to a channel.

An illustration of the priority present in the basic implementation and its effects on fairness is provided in the left-hand plot of Figure 3. The destination for traffic is node 8, and nodes 1 to 7 are the sources, each providing an equal offered load to the channel. One channel, channel 8, is simulated with the total normalized offered load being 150% (i.e., the channel is overloaded). In the figure, we plot cumulative traffic measurements from each source (snapshots) at regular intervals in time. As we move up the graph, time increases, and the total number of bits transmitted increases.

*This material is based upon work supported by the DARPA VLSI Photonics Program under grant DAAL01-98-C-0074.

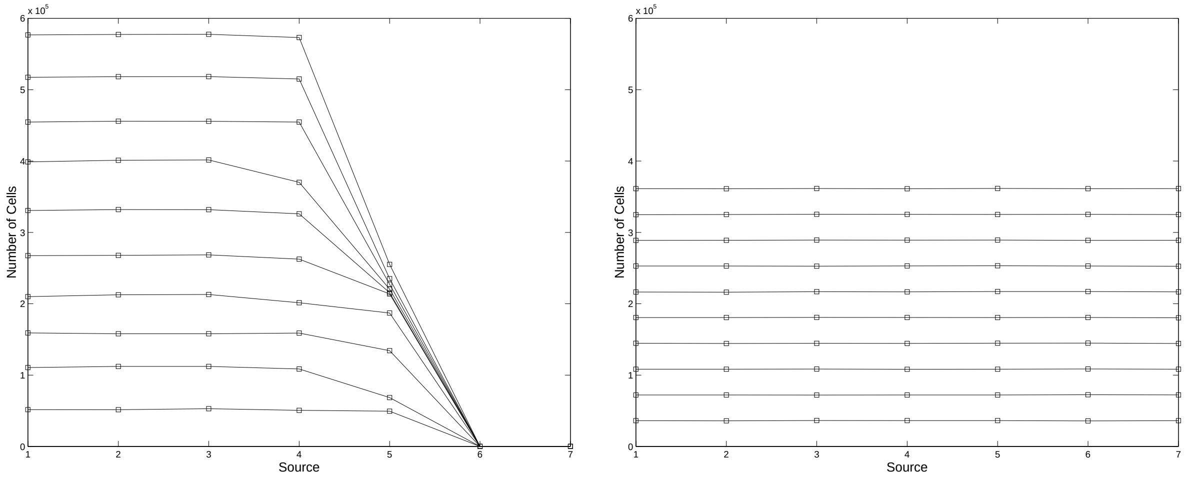

Figure 3: Cumulative traffic without DRR and with DRR.

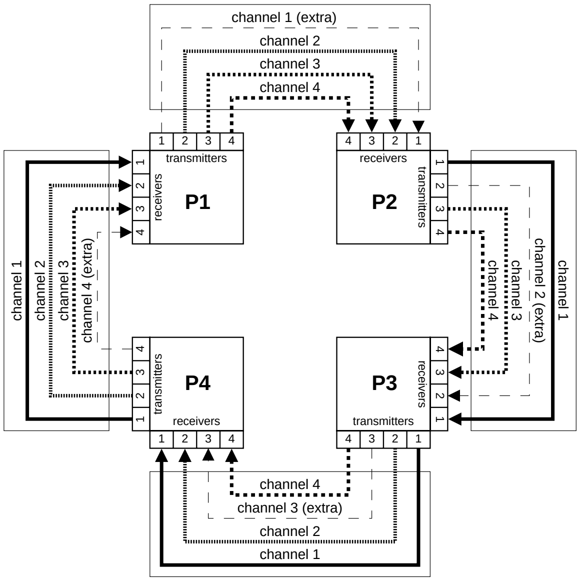

Figure 2: 4 node ring.

The flat lines for sources 1 to 4 indicate that these nodes are receiving roughly the same amount of access to the channel between snapshots. However, node 5 has only been able to deliver about half of the data that nodes 1 to 4 have delivered, and nodes 6 and 7 have been completely starved (i.e., no data delivered).

3 Fairness Protocol

The above example illustrates the need for an arbiter in the multiring architecture. While it is often advantageous for an arbiter to allow uniform access to all the nodes that contend for a channel, this may be application dependent. For example, it may be desirable to give a subset of the nodes more access to a channel (i.e., more bandwidth) than other nodes. An approach that suits the multiring is the Deficit Round Robin (DRR) scheduler.

DRR scheduling is described in [4] for use in internet switches and routers, where the contention is for an output link. The DRR scheduler has the following attractive properties:

- Flexibility. Nodes can be given different amounts of access to a channel by tuning parameters built into the protocol.

- Fast Decision Making. The DRR algorithm is fast since it needs to only examine the node in question to decide whether it should be given access.

- Fairness. DRR has been proven fair to the following extent: at any time, for equal priority channels, the difference in the amount of access granted to the most advantaged contender and the most disadvantaged contender is no more than three times the maximum message size.

Here, we adapt the DRR mechanism for use in a ring topology, where the information necessary to execute the algorithm is not all present at a single point in the system. Since every channel is attached to a particular receiver, we assign the DRR scheduler for that channel to its associated receiver. Prior to sending a message on channel j, a node sends a control signal to node j requesting access to the channel. When the DRR scheduling algorithm (executing on node j) decides that sender i should have access, it sends a control message to i granting access to the channel.

4 DRR Performance

The right-hand plot of Figure 3 illustrates the impact of the DRR scheduler. As before, we are simulating a total offered load of 150% channel capacity for 7 sources all contending to send data to node 8. The snapshot interval is identical to that of the previous plot. The only distinction between the two simulations is the presence of a DRR scheduler.

Unlike the earlier case, where downstream nodes (5, 6, and 7) received limited (or no) access to the channel, here each of the sources receives uniform access. The effective bandwidth delivered to each source is lowered (reflecting the total bandwidth capability of the channel), and it is fairly allocated to the contending sources.

Although not presented here, a similar effect occurs when the channel is used at less than 100% capacity. In this case, it is the latency experienced by each source that is impacted by the priority structure of the ring. We are currently running experiments to show that the DRR protocol balances the requests of the competing sources, providing each with equal access.

One of the features of embedded systems is that it is often necessary to guarantee application performance. This can be very difficult in a multiprocessor system if the interconnection network performance is unpredictable under certain circumstances. Providing fair access to communication channels can go a long ways towards enabling the performance of the system as a whole to be well understood and the operation of the system to be predictable.

References

[1] Ch'ng Shi Baw, Roger D. Chamberlain, and Mark A. Franklin. Design of an interconnection network using VLSI photonics and free-space optical technologies. In *Proc. of 6th Int'l Conf. on Parallel Interconnects*, pages 52–61, October 1999.

[2] M. Marsan et al. All-optical WDM multi-rings with differentiated QoS. *IEEE Communications Magazine*, pages 58–66, February 1999.

[3] D. Plant, J. Trezza, M. Venditti, E. Laprise, J. Faucher, K. Razavi, M. Chateauneuf, A. Kirk, and W. Luo. A 256 channel bi-directional optical interconnect using VCSELs and photodiodes on CMOS. In *Proc. of Optics in Computing*, June 2000.

[4] M. Shreedhar and G. Varghese. Efficient fair queueing using deficit round robin. In *Proc. of SIGCOMM*, pages 231–243, August 1995.