Optical interconnection networks with time slot routing

Ireneusz Szcześniak and Roman Wyrzykowski [a]

[a]Institute of Computer and Information Sciences
Częstochowa University of Technology
ul. Dąbrowskiego 73
42-200 Częstochowa
Poland

Abstract: In the article we describe the *time slot routing*, a routing scheme that requires only a simple design of interconnection networks. The simulation results show that the time slot routing performs optimally at the maximum uniform network load, and for uniform loads outperforms deflection routing.
Keywords: Interconnection networks, Beneš network, time slot, deflection, simulation.

1. Introduction

Data centers, computing centers and multiprocessor systems rely on interconnection networks which are electronic today and which can use optical networking for point-to-point communication only. However, optical switching is seen as a long-term candidate for interconnection networks since it can increase network throughput, and decrease latency and power consumption [1]. Optical interconnects are currently intensively researched as they promise to improve the interconnection networks in data centers which are growing and which are expected to grow dynamically [2].

A number of improvements have been recently proposed in the design of interconnection networks. In [3] bufferless routing is proposed, which is substantiated by extensive simulative performance evaluation. Small buffers and bufferless routing has been proposed a number of years ago [4].

In [5] authors on one hand recognize the bufferless advantage of deflection routing, and on the other hand notice the throughput degradation resulting from overloading the network, and so they propose methods of controlling congestion for bufferless deflection-routed networks.

We, in turn, introduce *time slot routing* for interconnection networks. Time slot routing is a form of time-division multiplexing (TDM) that is used in communication networks to share a transmission link, but we devised time slot routing as a simple and efficient means of routing packets in an interconnection network.

The article is organized as follows. In the next section, time slot routing is introduced. Then its performance is evaluated simulatively and compared to the performance of a network with deflection routing and with store-and-forward routing. The article ends with conclusions.

2. Time slot routing

The network interconnects n nodes which send and receive packets of constant duration synchronously, i.e. the according to a single clock. The network does not buffer packets and sends the packets from its inputs to outputs depending on the configuration of the switching fabric which changes every time slot. Nodes are responsible for buffering the packets they want to send, and for sending them at the required time slot, and so the network can be viewed as exhibiting the head-of-line blocking. The delay incurred by the optical interconnection network is small and equals to the delay it takes light to traverse the network. The delay is below a nanosecond, and we assume it equals to one time slot.

A permutation of the node numbers expresses the mapping of the inputs to the outputs of the interconnection network. Each time slot an interconnection network changes its state to implement a new predetermined permutation that depends on the time slot number t. There are $(n-1)$ permutations repeated periodically, since there are n nodes, and since a node should be able to send a packet to the other $(n-1)$ nodes. Three sample permutations p_1, p_2, and p_3 required for a 4×4 network are given in (1). For instance, permutation p_1 maps input 1 to output 2, 2 to 3, 3 to 4, and 4 to 1.

$$p_1 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \end{pmatrix}, p_2 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 1 & 2 \end{pmatrix}, p_3 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 1 & 2 & 3 \end{pmatrix} \quad (1)$$

Formally, assuming that permutation p_i allows source node s to send a packet to destination node $(s+i) \bmod n$, and that these $(n-1)$ permutations are repeated periodically, then destination node d of a packet sent from node s at time slot t is given by (2).

$$d = (s + 1 + t \bmod (n-1)) \bmod n \quad (2)$$

To demonstrate time slot routing scheme and to evaluate its performance, we chose the Beneš network for its small number of N required switching elements, as given by
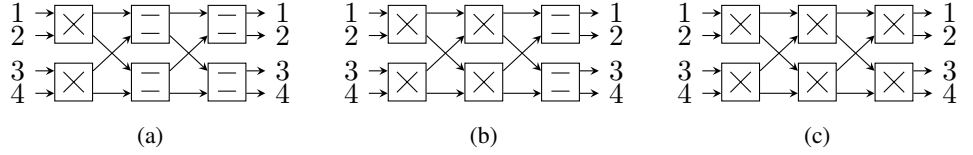
Fig. 1: Configurations of the 4×4 Beneš network.
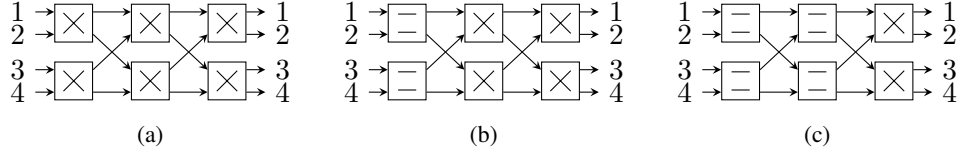

Fig. 2: Extra configurations of the 4×4 Beneš network.

(3), in comparison to other networks such as the Banyan networks. Any other network type can be used with time slot routing as long as it is rearrangeable; the nonblocking property is not required as the network can be rearranged every time slot. If the constant latency is required, then every path between the source and destination nodes should be of equal length, which holds true for the Beneš network.

$$N = \frac{n(2 \log_2(n) - 1)}{2} \tag{3}$$

Depending on the network type, a permutation can be implemented by more than one network configuration. Having different configurations for the same permutation can allow for less frequent reconfigurations of the switching elements, which would relax the requirements on the properties of switching elements, such as reconfiguration time, reducing the cost of the network and increasing the network performance.

For instance, Fig. 1 shows the configurations of the 4×4 Beneš network that implement permutations p_1', p_2', and p_3' given by (4), and Fig. 2 shows configurations that are different from the configurations in Fig. 1, but implement the same permutations. The sequence of configurations shown in Fig. 1a, 1b, 1c, 2a, 2b, 2c would produce the sequence of permutations p_1', p_2', p_3', p_3', p_2', and p_1', which would require the reconfiguration of every switching element at most every other time slot.

$$p_1' = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \end{pmatrix}, p_2' = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 2 & 1 \end{pmatrix}, p_3' = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 1 & 2 \end{pmatrix} \tag{4}$$

There are a number of advantages of time slot routing. First, time slot routing allows for a simple design of the interconnection network: it does not require buffers in the interconnection network, and neither packet header processing, even at the end nodes. Second, because time slot routing does not drop packets, there is no need for acknowledgments and retransmissions. Finally, there is no overhead incurred, since no packet header is used.

3. Performance evaluation

The performance of time slot routing is compared to the performance of deflection routing and store-and-forward routing for two sizes of the Beneš network: 16×16 (56 switching elements), and 64×64 (352 switching elements). The admission buffers at the nodes can grow without a limit, but the buffering at switching elements is subject to constraints as described below.

Deflection routing routes a packet along the shortest path, and misroutes it if the preferred output has already been taken by the other packet. If a packet prefers either output at a switching element (i.e. both outputs yield a path to the destination of the same length), one of them is chosen at random, which helps to improve the network performance and fairness. We used deflection routing results for comparison, because just like time slot routing, deflection routing does not require buffers.

In electronic interconnection networks, packets are usually sent with store-and-forward routing, which requires buffers at every switching element. We evaluated the performance of store-and-forward routing for three sizes of buffers: 1, 3 and 5, which are placed at the outputs of the 2×2 switching elements. When a packet arrives, it is enqueued in the buffer of its preferred output, i.e. the one that leads to the destination along the shortest path. If there is no space left in the buffer, the packet is dropped. We assume that a packet stays at a switching element at least one time slot due to the need of electronic processing and buffering.

We carry out the performance evaluation simulatively. The analysis of a TDM channel with a Poisson input stream should apply to time slot routing, but it is rather complex as the discrete and continuous time domains are intermixed [6], and so we reverted to simulations. Deflection routing could be evaluated analytically too, but its analysis is difficult and gives approximate results [7]. We evaluated the network throughput, the number of dropped packets, the admission queue delay, the total delay, and the admission queue size.

We assume that the network is uniformly loaded: each node sends packets to each of the other $(n - 1)$ nodes, and so there are $n(n - 1)$ flows in the network. The uniform load assumption is a common assumption for data center networks, and multi-threaded or distributed programs. At each source node there is one admission queue per flow that
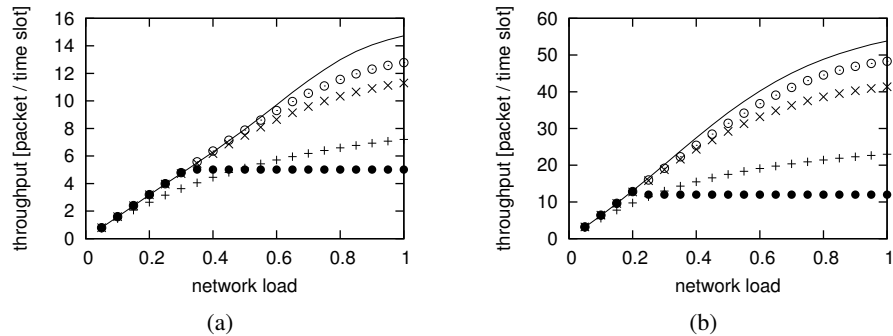
Fig. 3: Comparison of the mean network throughput.

we allow to grow without a limit. Packets arrive to the buffer according to the Poisson distribution with intensity $\lambda = l$ packets per time slot, where $l \in (0, 1)$ is the network load. The network is maximally loaded when $l = 1$, i.e. every node has on average one message to send every time slot. Since every node sends packets uniformly to other $(n - 1)$ nodes, the intensity of each flow is $\lambda/(n - 1)$.

3.1. Simulations

The performance evaluation was carried out with 2000 simulation runs. There were five groups of simulations: one for time slot routing, one for deflection routing, and three for store-and-forward routing for the buffer sizes of 1, 3, and 5. There were two network sizes ($n = 16, 64$) and twenty different loads ($l = 0.05, 0.1, \ldots, 1.0$), and so there were $5 \times 2 \times 20 = 200$ test cases, where each test case was simulated ten times with different pseudo-random values, totaling 2000 simulation runs.

We calculated the mean values of the metrics of interest for a test case based on the results of ten simulation runs for that test case. We also calculated the standard errors for the mean values, and found that they were below 1% of the mean value.

3.2. Comparison

Figures 3, 4, 5, 6, and 7 show the simulation results. In each of the figures, the subfigures (a) are for the 16×16 network, and (b) for the 64×64 network. In each of the figures the results for time slot routing are shown with the solid lines, and for deflection routing with the bullets (●). The results for store-and-forward routing are shown with the pluses (+) for buffers of size 1, with the crosses (×) for the buffers of size 3, and with the circles (○) for the buffers of size 5. There are no error bars shown for the data points, because the standard errors were too small to be drawn.
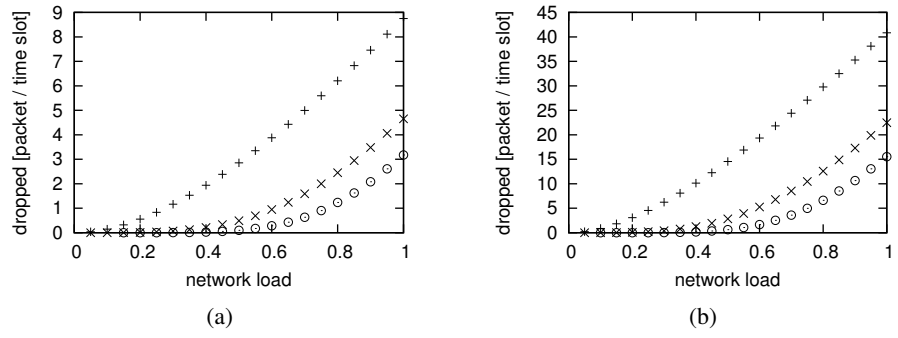
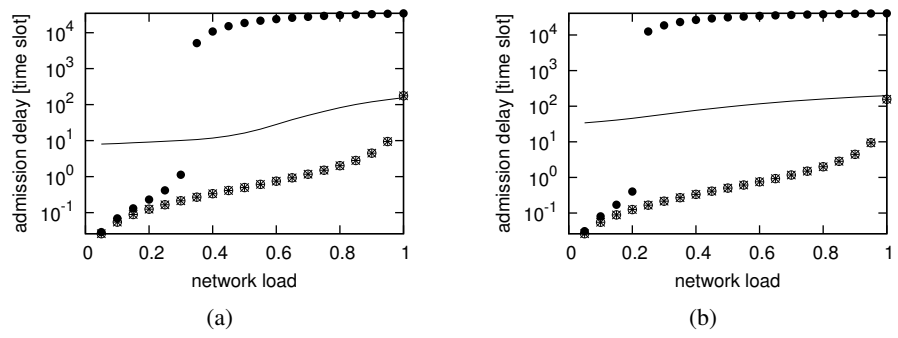Fig. 4: Comparison of the mean number of dropped packets.

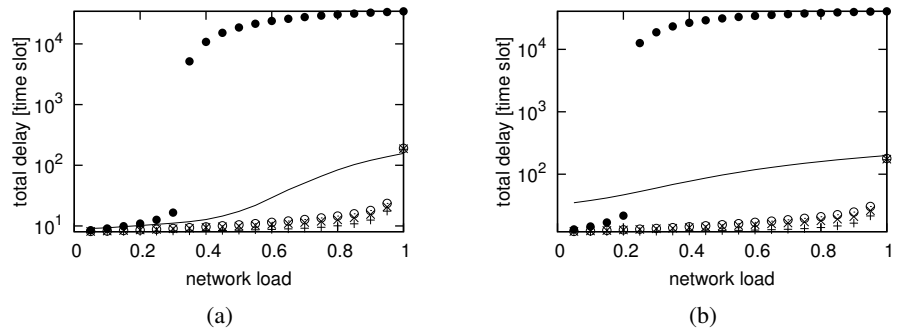Fig. 5: Comparison of the mean admission delay.

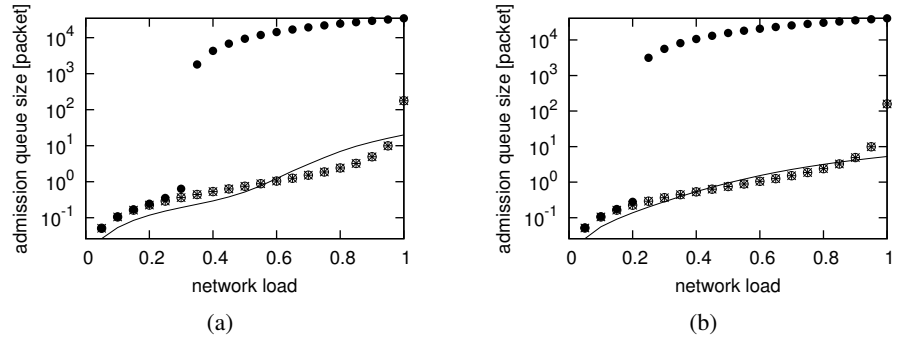Fig. 6: Comparison of the mean total network delay.

Fig. 7: Comparison of the mean admission queue size.

The mean network throughput is shown in Fig. 3. Deflection routing performs poorly for high loads, because deflected packets saturate the network, thwarting packet admission and delivery. In contrast, time slot routing performs optimally in that it delivers all the offered load, reaching the maximum throughput at heavy loads: at most n packets per time slot for the $n \times n$ network. Also store-and-forward routing with buffers of size 5 performs worse than time slot routing.

The mean number of packets dropped per time slot is shown in Fig. 4. Since time slot routing and deflection routing do not lose packets, their results are not shown in the figure. Store-and-forward routing, however, suffers a large number of dropped packets, even when buffers of size 5 are used.

The mean admission delay measured in time slots is shown in Fig. 5. Generally, deflection routing causes the largest delay because deflected packets live in the network long, blocking admission of new packets. The mean admission delay for time slot routing is at least $(n-1)/2$, and so it is larger than the admission delay for store-and-forward routing [6].

The mean total delay measured in time slots is shown in Fig. 6, which is the sum of the number of time slots spent waiting in the admission queue, and the number of time slots spent in the network. Again, time slot routing outperforms deflection routing. For and store-and-forward routing, the admission delay dominates, and so the time slot routing in general performs worse than store-and-forward routing.

The mean size of admission queues is shown in Fig. 7. For deflection routing, the long-lived deflected packets cause the admission queues to grow to very large sizes. Time slot routing and store-and-forward routing cause the admission queues to grow to comparable sizes.

4. Conclusion

In the article we described time slot routing, and reported the simulative performance evaluation results for the network of the Beneš type. We showed that time slot routing outperforms deflection routing, and compares favorably with store-and-forward routing. Time slot routing, unlike store-and-forward routing, allows for a simple network design without buffers and header processing.

An interesting problem for future work is finding a network type, or algorithms of configuring existing network types, that would require only infrequent reconfiguration of switching elements.

References

[1] "International technology roadmap for semiconductors: interconnect chapter," 2011.

[2] C. Kachris and I. Tomkos, "A survey on optical interconnects for data centers," *IEEE Communications Surveys Tutorials*, vol. 14, no. 4, pp. 1021–1036, 2012.

[3] T. Moscibroda and O. Mutlu, "A case for bufferless routing in on-chip networks," *SIGARCH Comput. Archit. News*, vol. 37, pp. 196–207, June 2009.

[4] M. Enachescu, Y. Ganjali, A. Goel, N. McKeown, and T. Roughgarden, "Routers with very small buffers," *Proceedings of INFOCOM 2006*, pp. 1–11, April 2006.

[5] G. Nychis, C. Fallin, T. Moscibroda, and O. Mutlu, "Next generation on-chip networks: what kind of congestion control do we need?," in *Proceedings of the Ninth ACM SIGCOMM Workshop on Hot Topics in Networks*, Hotnets '10, (New York, NY, USA), pp. 12:1–12:6, ACM, 2010.

[6] K.-T. Ko and B. Davis, "Delay analysis for a TDMA channel with contiguous output and poisson message arrival," *IEEE Transactions on Communications*, vol. 32, pp. 707–709, June 1984.

[7] I. Szcześniak, B. Mukherjee, and T. Czachórski, "Approximate analytical performance evaluation of synchronous bufferless optical packet-switched networks," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 3, pp. 806–815, October 2011.