

Up-Down Routing Based Deadlock Free Dynamic Reconfiguration in High Speed Local Area Networks

Naresh Kumar¹, Renu Vig² and Deepak Bagai³

¹ Kurukshetra University

Received: 27 February 2011 Accepted: 22 March 2011 Published: 6 April 2011

Abstract

Dynamic reconfiguration of high speed switched network is the process of changing from one routing function to another while the network remains in running mode. Current distributed switch-based interconnected systems require high performance, reliability and availability. These systems changes their topologies due to hot expansion of components, link or node activation and deactivation. Therefore, in order to support hard real-time and distributed multimedia applications over a high speed network we need to avoid discarding packets when the topology changes. Thus, a dynamic reconfiguration algorithm updates the routing tables of these interconnected switches according to new changed topology without stopping the traffic. Here, we propose an improved deadlock-free partial progressive reconfiguration (PPR) technique based on UP/DOWN routing algorithm that assigns the directions to various links of high-speed switched networks based on pre-order traversal of computed spanning tree. This improved technique gives better performance as compared to traditional PPR by minimizing the path length of packets to be transmitted. Moreover, the proposed reconfiguration strategy makes the optimize use of all operational links and reduces the traffic congestion in the network. The simulated results are compared with traditional PPR.

Index terms—

routing tables when there is change in network topologies due to network components failures and addition of new one. In such cases a dynamic reconfiguration algorithm analyze the new network topology and updates the routing table by replacing the old routing function with the new updated routing function. Although both routing functions are deadlockfree still it may create deadlock during reconfiguration because one of the routing function may take turns that are not allowed in the other routing function by making a cycle in the network.

In the literature, static reconfiguration (designed for Autonet [13,15] and ??yrinet[2,8] networks) replaces the routing function from old to new by stopping the network traffic. Hence it creates negative impact on the network service availability and the performance of the overall network degrades.

Many new schemes have been proposed recently to enhance network service availability while the network change over from one routing function to another. These new schemes updates the routing table of switched network during run time when there is any change in the network topologies due to addition/removal of links/nodes and are known as dynamic reconfiguration techniques.(Partial Progressive Reconfiguration (PPR)[3], Skyline [6], Double scheme [10], and Simple Reconfiguration [7]), NetRec [16] and Dynamically scaling Algorithm(DSA) [17]. These schemes are designed for the networks that uses distributed routing. Double Scheme requires extra resources like virtual channels for deadlock handling that occur during transition of old routing function to new routing function. Simple reconfiguration requires a special packet called token to avoid deadlock. In this scheme firstly all packets are transmitted by old routing function, then token and finally the packet transmission is based on new routing function. NetRec was proposed as a dynamic reconfiguration scheme to increase the network availability in the presence of a permanent node fault. It restores the network connectivity

by building a tree that spans all immediate neighbours of the faulty node that are still connected to network. The NetRec Scheme [16] requires every switch to maintain information about nodes some number of hops away and is only applicable to wormhole networks. NetRec is

1 INTRODUCTION

2 H

extended to dynamically reconfigure the network for the case of newly joining nodes called Dynamically Scaling Algorithm (DSA). The solution in DSA is based on performing sequence of partial routing table updates, while dropping the user messages in all selected nodes until the restoration is completed. PPR requires a sequence of synchronizing steps to progressively update old forwarding table entries to new ones while ensuring that no cycles form. The PPR [3] approach correct an invalid UP/DOWN graph after change in network topology. Double Scheme [10] uses the concept of virtual channels to avoid deadlock while reconfiguring the network. Simple Reconfiguration [7] uses a packet called a token to avoid deadlock by ensuring that the packet which belongs to old routing function are transmitted first, then the token, and finally packet transmission is based on the new routing function. A management mechanism [11,12] and zeroconfiguration hierarchical UP/DOWN routing[19] has been discussed for distributed calculation of the new routing function when there is any change in the switched-network topology. In this mechanism the distributed path-computation gives better performance when compared to existing centralized pathcomputation, still the routing function was updated statically, therefore degrades the performance.

In this research work, we propose to apply a new and very efficient dynamic reconfiguration strategy based on PPR that makes transformation of an invalid UP/DOWN graph (that is due to change in topology) into a valid UP/DOWN graph, while ensuring that there is no cycles in the graph to make it deadlock-free. Moreover, our dynamic reconfiguration does not use any additional resources such as virtual channels or special packet a token. This new proposed scheme gives better performance than PPR by distributing the traffic through optimize use of all links and reducing the congestion on root node of the spanning tree.

The next section discuss the concepts of UP/DOWN routing based PPR scheme and provides background information on previous studies of reconfiguration of UP/DOWN routing networks. Section 3 gives the details of improved dynamic reconfiguration strategy and assigning the directions UP/DOWN to the operational links based on pre-order traversal of spanning tree. In section 4, the performance of the proposed reconfiguration strategy is evaluated. Finally, section 5 concludes this research work and proposes some future work.

PPR scheme is based on deadlock-free UP/DOWN routing algorithm [14] for irregular network topologies. The UP/DOWN routing algorithm is based on a cycle-free assignment of directions to the operational links in the network. For each link in the network, one direction is named up and the other is down. Deadlock avoidance is achieved by using legal routes. A packet never use a link in the up direction after having used one in the down direction. Messages can traverse zero or more links in the up direction, followed by zero or more links in the down direction. Therefore, deadlocks are prevented and cycles in the channel dependency graph [4] are avoided. Major problem of old PPR is the random assignment of UP/DOWN direction to links between two or more nodes at the same label.

A sink node [3] does not have outgoing uplinks(a node that is not the source of any link). There are no legal routes between any two sink nodes due to the restrictions imposed by UP/DOWN routing algorithm because each route would require a down to up transition, which is not allowed. So there is always a single sink node in directed acyclic graph based on UP/DOWN routing algorithm. A break node is the source of two or more links. This node breaks the cycles formation to avoid deadlock in directed graph. The graph of Fig 1 includes various break nodes like node j, e, d. A correct graph contains a single sink node for UP/DOWN routing. It includes as many break nodes as necessary to break all cycles for deadlock freedom. The graph shown in Fig 1 is a correct graph. On the other hand, an incorrect graph does not meet the restrictions imposed in the correct graph. An incorrect graph has the absence of a sink node, presence of more than one sink node, or there are cycles in the graph. Fig. 2 tells about an incorrect graph with two sink nodes ?b' and ?c' after removal of node ?a' from Fig. ?? correct graph. A PPR scheme changes the direction of operational links to give a correct graph from an incorrect graph as shown in Fig. ??In the literature, several algorithms have been proposed for constructing an UP/Down directed graph. Traditional proposals are based on the computation of a spanning tree which is built using a breadth-first search (BFS) [14], a depth-first search(DFS) [13], or a propagation-order spanning tree(POST) [15].

3 Network

topology changes due to addition/removal of switches/ links, then our dynamic reconfiguration scheme calculates a new routing function which ensures that packets that belong to the new routing function can not take turns in the old routing function, and vice versa. Therefore, packets of old and new routing functions can unrestrictedly coexist in the network without creating deadlocks. This improved PPR scheme is based on the concept of assigning the UP/DOWN directions to operational links by pre-order traversal of spanning tree in which a unique label is for each node of the graph. We also present lemmas to support that, when an UP/DOWN graph for the new

103 topology is designed based on pre-order traversal of spanning tree, the routing function is updated without the
104 risk of transient deadlocks.

105 **4 Definition 1**

106 Assume that two UP/DOWN directed graphs, G1 and G2, represented two network topologies. Then G1 and
107 G2 are corrected.

Fig 4 presents UP/DOWN graph G1 which is correct. Fig. 5 shows again a correct graph G2 after removal of root node 'a' of graph G1 in Fig. 4 Lemma1

110 Assume that an UP/DOWN directed graph G1 based on pre-order traversal of spanning tree is correct. Then,
111 it is always possible to obtain a correct graph G2 from G1 when any node or link is added or removed.

112 5 Proof

In a correct UP/Down graph, each possible cycle must contain at least one node with two incoming up-links and
at least one node with two outgoing uplinks. a b c d e f g h i j k a [x] [0] [1] [0,1] [0,1] [0] [1] [1] [1] [0,1] [1] b
[0] [x] [0] [1] [2,0] [3] [0] [0] [0] [1,2,3,0] [0] c [0] [0] [x] [1,0] [2,0] [0] [3] [4] [5] [1,2, 3,4,0] [3,5] d [0,1] [0,1] [1,0] [x]
[0,1] [0] [1,0] [1,0] [1,0] [2,1,0] [1,0] e [0,2] [0,2] [2,0] [0,2] [x] [0] [2,0] [2,0] [2,0] [1,2,0] [2,0] f [0] [0] [0] [0] [0] [x] [0]
[0] [0] [1,0] [0] g [0] [0] [0] [0] [0] [0] [x] [0] [1,0] [2,0] [0] h [0] [0] [0] [0] [0] [0] [0] [x] [0] [1,0] [0] i [0] [0] [0] [0] [0]
[0] [0] [0] [x] [0] [1 ,0] j [0,1,2, 3,4] [0,1,2, 3,4] [1,3,4 ,2,0] [0,1, 2,3,4] [1,2,0 ,3,4] [2,1, 0,3,4] [3,4, 2,1,0] [4,3,2, 1,0]
[4,3,2 ,1,0] [x] [4,3,2 ,1,0] k [0,1] [0,1] [0,1] [0,1] [0,1] [0,1] [0,1] [0,1] [1,0] [0,1] [x] b c d e f g h i j k b [x] [2] [1] [2]
[3] [1,2] [1,2] [1,2] [1,2,3] [1,2] c [1,2] [x] [1,2] [2,1] [1,2] [3] [4] [5] [3,4, 1,2] [3,5] d [0] [1,0] [x] [0] [0] [1,0] [1,0] [1,0] [1,0]
[2,1,0] [1,0] e [0] [2,0] [0] [x] [0] [2,0] [2,0] [2,0] [1,2,0] [2,0] f [0] [0] [0] [x] [0] [0] [0] [0] [0] [0] [0] [0] [x]
[0] [0] [1,0] [2,0] h [0] [0] [0] [0] [0] [0] [x] [0] [1,0] [0] i [0] [0] [0] [0] [0] [0] [0] [0] [0] [1,0] j [0,1,2 ,3,4] [0,1, 3,4,2]
[0,1,2 ,3,4] [1,0,2 ,3,4] [2,0, 1,3,4] [3,0, 1,2,4] [4,0, 1,2,3] [3,4 ,0,1,2] [x] [3,0, 1,2,4] k [0,1] [0,1] [0,1] [0,1] [0,1] [0,1]
[0,1] [1,0] [1,0] [x] j k a [x] [0] [1,0] [0] [0] [0] [1,0] [1,0] [1,0] [0] [1,0] b [0] [x] [0,2,1] [1] [2,1] [3,1] [0,1] [0,1] [0,1] [1]
[0,1] c [0,1,2] [0,1] [x] [1,0,2] [2,0,1] [0,1,2] [3,2, 1,0] [4,2, 1,0] [5,3, 2,1,0] [2,1,0] [3,2, 1,0] d [0] [0] [1,0,2] [x] [0,2]
[0,2] [1,2,0] [1,2,0] [1,0,2] [2] [1,2,0] e [0,1] [0,1] [2,0] [0,1] [x] [0,1] [1,2,0] [1,2,0] [2,0,1] [1,0] [1,2,0] f [0,1] [0,1] [0,1]
[0,1] [0,1] [x] [1,0] [1,0] [1,0] [1,0] [1,0] g [0,1] [0,1] [0,1] [1,0] [1,0] [1,0] [x] [0,1] [0,1,2] [1,0] [2] h [0,1] [0,1] [0,1] [1,0]
[1,0] [1,0] [1,0] [x] [0,1] [1,0] [1] i [0,1] [0,1] [0,1] [0,1] [0,1] [0,1] [0,1] [0,1] [x] [0,1] [1,0] j [0] [0] [0] [0] [1,0] [2,0] [3,0]
[4,0] [3,1,0] [x] [3,1,0] k [0] [0] [0] [0] [0] [0] [0] [0] [0] [0] [1,0] [0] [x] i Switch b c d e f g h i j k b [x] [1,2] [1] [2,1] [3,1]
[1,2] [1,2] [1,2] [1] [1,2] c [1,2] [x] [1,2] [2,1] [2,1] [3,2,1] [4,2,1] [5,2,1] [1,2] [3,2,1] d [0] [1,2,0] [x] [0,2] [0,2] [2,1,0]
[2,1,0] [1,0,2] [2] [1,2,0] e [0,1] [2,1,0] [0,1] [x] [0,1] [2,1,0] [2,1,0] [2,1,0] [1,0] [2,1,0] f [0,1] [0,1] [0,1] [0,1] [x] [1,0]
[1,0] [1,0] [1,0] [1,0] g [0,1] [0,1] [0,1] [0,1] [1,0] [x] [0,1] [0,2,1] [1,0] [2,1,0] h [0,1] [0,1] [0,1] [0,1] [1,0] [0,1] [x] [0,1]
[1,0] [0,1] i [0,1] [0,1] [0,1] [0,1] [0,1] [0,1] [x] [1,0] [1,0] j [0] [0] [1,0] [0] [1,0] [2,0] [3,0] [4,0] [3,1,0] [x] [3,1,0] k [0]
[0] [0] [0] [0] [0] [0] [1,0] [0] [x] j k l m b [X] [1,2] [1] [2,1] [3,1] [1,2] [1,2] [1,2] [1] [1,2] [1] [1,2] c [1,2] [X] [5] [3,1,0]
k [0] [0] [0] [0] [0] [0] [0] [0] [1,0] [0] [X] [0] [2] l [0] [0] [0] [0] [0] [0] [0] [0] [0] [0] [0] [0] [X] m [0] [0] [0] [0] [0] [0] [0]
[0] [0] [X]

cycle in G2, we have three options with respect to break node placement. Thus, it is always possible to construct a correct graph G2 which is also like G1. Assume that two up/down directed graphs G1 and G2 where G2 is also correct after removal of any node or link of correct graph G1. Assume that G2 is a subgraph of G1 in which each link that connects a break node in G1 with the corresponding break node in G2 is suppressed, and that G2 is a correct UP/DOWN graph. Then, according to lemma 1, we can guarantee that G2 is also a correct UP/DOWN graph. Thus it possible to define a fully connected deadlock-free routing function R2 over G2. R2 satisfies the routing-restrictions of both G1 and G2 since all the break nodes either have the same locations or have been removed.

145 Preorder Traversal Based UP/DOWN direction assignment to various link. 1) Computing a spanning tree.

2) Preorder traversal of spanning tree and assign labeling to nodes through horizontal traversal based distance
of each node from root node. 3) Next assigning UP/DOWN directions to various links likewise previous strategy.

148 4) Finally, updation of routing tables

149 6 Switch Deactivation

150 The deactivation of switch including the root node produces a new root node. Switch deactivations imply that
151 messages routed to remove components must be discarded. In this case, shorter reconfiguration time implies less
152 discarded messages which is the characteristic of this improved PPR scheme.

153 7 Switch Activation

154 When a new switch is added, a direction must be assigned to the links connecting to it in such a way that the
155 down direction goes toward the new switch and it should no produce cycles in the directed graph.

In this section, we evaluate the performance of the improved PPR algorithm proposed in section 3. Our dynamic reconfiguration scheme is compared to traditional PPR scheme. New mechanism requires very less computations time for updation of routing function from old one to new, when there is change in network topology due to addition or removal of switch nodes or links.

8 a) Switch Model

A switch consist of crossbar, a routing and configuration unit, and many full-duplex links. The routing and configuration unit provides the output channel for multiple packets to cross a switch. Table look-up routing is used. Each input channel has 2 set of buffers: user and control buffers. Control buffers handle control messages generated in each reconfiguration process when there is any change in network topology.

We have assumed that one clock cycle is required to access the routing table and provide the output link for the message. The high-speed switched networks are consist of a set of switches interconnected through point-to-point links and hosts are connected to those switches through a network interface card in a irregular fashion. We have evaluated the performance of different sizes of networks. Several different networks with irregular topologies are considered in order to perform a detailed study. The irregular topologies have been randomly generated.

9 c) Message Generation

For each simulation run, we have considered that message generation rate is constant and the same for all the nodes. Once the network has reached a steady state, the flit generation rate is equal to the flit reception rate(traffic).We have evaluated the full range of traffic, from low load to saturation. On the other hand, we have considered the message destination is randomly chosen among all the nodes d) Simulation Results

In this section, we show the simulation results for improved PPR scheme based on pre-order traversal of spanning tree. The simulation results are compared with existing PPR scheme. The simulation used for this work is developed with the IRFlexiSim Simulator [18]. First of all, we have discussed the way in which the path computation time is reduced by using improved PPR strategy in section 3. In order to increase the accuracy of the results, each experiment is repeated several times. The numbers of simulation runs for each topology are presented graphically. The number of data packets that are discarded during the topology change assimilation process gives an indication of the level of service a network can provide to applications. Fig. ?? compares the amount of packets that are discarded for PPR and improved PPR scheme. The results in Fig ?? shows that, for a switch removal the rate at which data packets are discarded is notably lower for new PPR scheme. The main reason is that, for improved PPR scheme minimum changes are required to update the routing tables. In case of switch additions, no packets are discarded due to inactive ports because no old roots include the switch that was just added. There is no packet discarding when a switch addition occur because it does not destroy any of the existing paths. To conclude this evaluation, Figures as shown below illustrates the instantaneous behavior of both the old PPR and new PPR scheme. The improved PPR scheme reduces the number of packets discarded during reconfiguration and increases the performance by distributing the traffic over all links of the network. Thus, it reduces the latency of packets and increases throughput by optimize use of all links and lowering the congestion at root node as shown in Fig ??.

Use In this paper, we have proposed and evaluated a simple improved reconfiguration strategy to compute the UP/DOWN routing tables. The improved methodology makes use of pre-order traversal of spanning tree in order to assign UP and DOWN directions to links is able to impose less routing restrictions to remove the cyclic channel dependencies in the network than those imposed when using the paths and a better traffic balance, thus increasing channel utilization in the network .At a minor computational cost, the new routing function is designed to ensure that packets routed according to the old and the new routing functions can unrestrictedly coexist in the network, without the risk of forming deadlocks. Simulation results show that this significantly reduces the amount of packets that are discarded during the topology change assimilation. Moreover, our strategy does not required additional resources such as virtual channels, and it could easily be implemented in current commercial systems. As future work, we plan to extend the proposed methods in order to support other routing algorithms in addition to UP/DOWN.

¹May Up-Down Routing Based Deadlock Free Dynamic Reconfiguration in High Speed Local Area Networks ©2011 Global Journals Inc. (US)

²May Up-Down Routing Based Deadlock Free Dynamic Reconfiguration in High Speed Local Area Networks ©2011 Global Journals Inc. (US)

³May Up-Down Routing Based Deadlock Free Dynamic Reconfiguration in High Speed Local Area Networks ©2011 Global Journals Inc. (US)

⁴May Up-Down Routing Based Deadlock Free Dynamic Reconfiguration in High Speed Local Area Networks ©2011 Global Journals Inc. (US)

⁵May Up-Down Routing Based Deadlock Free Dynamic Reconfiguration in High Speed Local Area Networks ©2011 Global Journals Inc. (US) Lemma 2

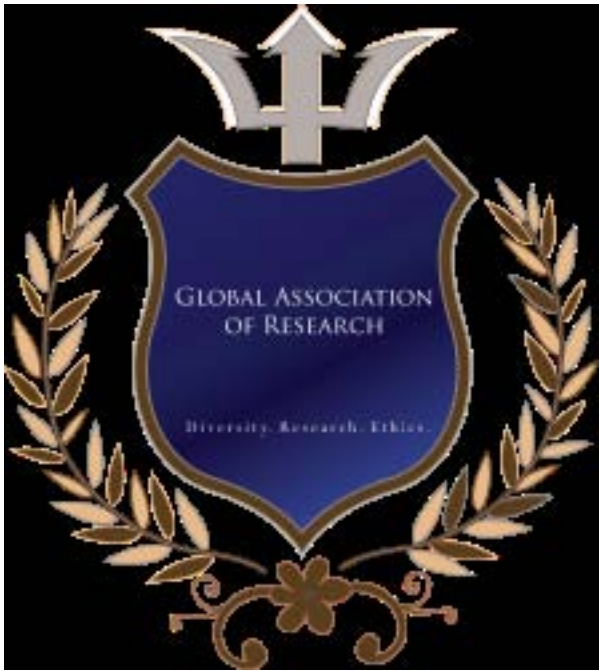
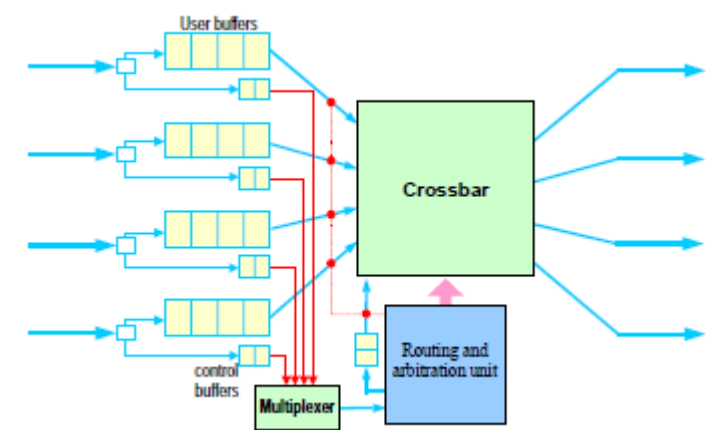


Figure 1: Fig1:



2

Figure 2: Fig. 2

1

Switch

Figure 3: Table 1 :

3

Switch a	b	c	d	e	f	g	h	i
----------	---	---	---	---	---	---	---	---

Figure 4: Table 3

5

Figure 6: Table 5

1. Advanced Switching Interconnect Special Interest Group, Advanced Switching Core Architecture Specifics

Figure 7:

-
- [Lbanez et al. ()] , Guillermo Lbanez , Alberto Garcia -Martinez , Juan A Carral , Pedro A Gonzalez , Arturo Azcorra , Jose M Arco . HURP/HURBA: Zero-configuration. <http://www.ceng.usc.edu/smart/tools.html> 2010.
- [Avresky et al. (2004)] ‘?Dynamically Scaling System Area Networks?’. D R Avresky , Y Varoglu , M Marinov . *IEEE Proceedings of the 12 th EuroMicro Conference On Parallel , Distributed and Network-Based Processing*, Feb 2004. p. .
- [Robles -Gómez et al. ()] ‘A Model for the Development of ASI Fabric Management References Références Referencias Protocols’. A Robles -Gómez , E M García , A Bermúdez , R Casado , F J Quiles . *LNCS* Nagel, W.E., Walter, W.V., Lehner, W. (ed.) 2006. 2006. Springer. 4128.
- [Sancho et al. (2000)] ‘A new methodology to compute deadlock-free routing tables for irregular networks’. J C Sancho , A Robles , J Duato . *Proc. the 4th Workshop on Communication, Architecture, and Applications for Network-based Parallel Computing*, (the 4th Workshop on Communication, Architecture, and Applications for Network-based Parallel Computing) January 2000.
- [Robles-Gómez et al. ()] ‘A proposal for managing ASI fabrics’. A Robles-Gómez , A Bermúdez , R Casado , F J Quiles , T Skeie , J Duato . *Journal of System Architecture (JSA)* 2008.
- [Rodeheffer and Schroeder (1991)] ‘Automatic reconfiguration in Autonet’. T L Rodeheffer , M D Schroeder . *SRC Research Report 77 of the ACM Symposium on Operating Systems Principles*, October 1991.
- [Schroeder et al. (1991)] ‘Autonet: a high-speed, self-configuring local area network using point-to-point links’. M D Schroeder , A D Birrell , M Burrows , H Murray , R M Needham , T L Rodeheffer , E H Satterthwaite , C P Thacker . *IEEE Journal on Selected Areas in Communications* October 1991. 9 (8) .
- [Dally and Seitz ()] ‘Deadlock-free message routing in multiprocessor interconnection networks’. W J Dally , C L Seitz . *IEEE Transactions on Computers* 1987. 36 (5) .
- [Pinkston et al. (2003)] ‘Deadlockfree dynamic reconfiguration schemes for increased network dependability’. T M Pinkston , R Pang , J Duato . *IEEE Transactions on Parallel and Distributed Systems* June 2003. 14 (6) .
- [Netrec and Avresky ()] *Dependable Network Computing*, ; Netrec , Dimiter Avresky . 2000. Dordrecht: Kluwer Academic Publishers. p. 10.
- [Lysne and Duato (2000)] ‘Fast dynamic reconfiguration in Irregular networks’. O Lysne , J Duato . *Proc. Int. Conference on Parallel Processing*, (Int. Conference on Parallel essing) August 2000.
- [Myrinet ()] Inc Myrinet . <http://www.myri.com/> *Guide to Myrinet-, Switches and Switch Networks*, 2000.
- [Lysne et al. (2004)] ‘Simple deadlockfree dynamic network reconfiguration’. O Lysne , J M Montañana , T M Pinkston , J Duato , T Skeie , J Flich . *Proc. of the International Conference on High Performance Computing*, (of the International Conference on High Performance Computing) December 2004.
- [Horst (1995)] ‘Tnet: A reliable System area network’. R Horst . *IEEE Micro* Feb. 1995. p. .