

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Motivation and Objectives . . . . .	6
1.2	Contributions . . . . .	7
1.3	Organization . . . . .	8
<b>2</b>	<b>Interconnection Networks for HPC Clusters</b>	<b>9</b>
2.1	Network Design for Clusters . . . . .	10
2.1.1	Data Units . . . . .	11
2.1.2	Switching . . . . .	12
2.1.3	Virtual Channels . . . . .	13
2.1.4	Flow Control . . . . .	14
2.1.5	Arbitration . . . . .	16
2.1.6	Topology . . . . .	17
2.1.7	Routing Algorithm . . . . .	20
2.1.7.1	Taxonomy . . . . .	20
2.1.7.2	The Deadlock Problem . . . . .	22
2.1.7.3	The Livelock Problem . . . . .	24
2.1.7.4	Routing Algorithms for Clusters . . . . .	25
2.1.8	Network Reconfiguration . . . . .	27
2.1.8.1	Static and Dynamic Reconfiguration . . . . .	28
2.1.8.2	Existing Proposals on Dynamic Reconfiguration . . . . .	28
2.2	InfiniBand Networks . . . . .	29
2.2.1	InfiniBand Architecture . . . . .	31
2.2.2	Virtual Channels and Flow Control . . . . .	33
2.2.3	Routing in IBA . . . . .	33
2.2.4	Subnet Manager . . . . .	36

<b>3</b>	<b>Fault Tolerance for Interconnection Networks</b>	<b>37</b>
3.1	Basic Concepts . . . . .	37
3.2	Fault Models . . . . .	40
3.3	Fault-Tolerance Mechanisms . . . . .	41
3.3.1	Fault-Tolerant Routing Algorithms . . . . .	42
3.3.1.1	Hardware-Based Fault Tolerant Routing Algorithms	42
3.3.1.2	Software-Based Fault Tolerant Routing Algorithms	46
3.4	Applying Fault Tolerant Mechanisms to PC Clusters . . . . .	47
3.4.1	Fault Tolerance in IBA . . . . .	47
3.4.2	Contributions of the Thesis . . . . .	48
<b>4</b>	<b>Fault-Tolerant Routing Methodologies Based on Disjoint Paths</b>	<b>51</b>
4.1	Preliminaries . . . . .	51
4.2	TFTR Methodology . . . . .	55
4.2.1	TFTR Stages . . . . .	56
4.2.2	Enhancements to the TFTR Methodology . . . . .	63
4.2.3	Computational Cost . . . . .	66
4.3	SPFTR Methodology . . . . .	67
4.3.1	Motivation . . . . .	67
4.3.2	SPFTR stages . . . . .	69
4.3.3	Routing Algorithm . . . . .	70
4.3.4	Route Patterns and Network Regions . . . . .	71
4.3.5	SLtoVL Table Initialization . . . . .	73
4.3.6	Extending the Methodology . . . . .	78
4.3.7	Optimizing the methodology: A-SPFTR . . . . .	81
<b>5</b>	<b>Fault-Tolerant Routing Methodology Based on Reachability</b>	<b>83</b>
5.1	Motivation . . . . .	83
5.1.1	RFTR Preliminaries . . . . .	84
5.2	RFTR Description . . . . .	86
5.2.1	Direct and Indirect Reachability . . . . .	86
5.2.2	RFTR Methodology . . . . .	87
5.3	RFTR Complexity . . . . .	91
5.4	RFTR on InfiniBand . . . . .	92
<b>6</b>	<b>Fault-Tolerance Methodology Based on Reconfiguration</b>	<b>95</b>
6.1	Motivation . . . . .	95
6.2	Basic EBR Description . . . . .	97

6.3	Triggering the EBR Mechanism . . . . .	98
6.4	Routing Path Computation . . . . .	99
6.5	Path Distribution . . . . .	100
6.6	Deadlock Recovery Process . . . . .	101
6.7	Overlapping Multiple Reconfiguration Processes . . . . .	104
<b>7</b>	<b>Evaluation Model</b>	<b>107</b>
7.1	Introduction to Simulation Modeling . . . . .	107
7.1.1	Simulation Tool . . . . .	108
7.1.2	Advantages and Disadvantages . . . . .	109
7.2	Network Model . . . . .	110
7.2.1	Topologies . . . . .	110
7.2.2	End to End Flow Control . . . . .	111
7.2.3	Traffic Patterns . . . . .	111
7.2.4	Simulation Parameters . . . . .	112
7.3	Modelling and Analyzing Faults . . . . .	113
7.4	Compared Reconfiguration Mechanisms . . . . .	114
7.5	Collecting Results . . . . .	116
7.5.1	Performance Metrics . . . . .	118
<b>8</b>	<b>Evaluation</b>	<b>121</b>
8.1	TFTR Methodology . . . . .	121
8.1.1	Number of Paths . . . . .	122
8.1.2	Singular Cases . . . . .	122
8.1.3	Resources Required . . . . .	124
8.1.4	Quality of Routing Paths . . . . .	125
8.1.5	TFTR with Additional Virtual Channels . . . . .	126
8.1.6	Network Throughput . . . . .	126
8.1.7	Computational Cost . . . . .	128
8.1.8	Summary . . . . .	128
8.2	SPFTR and A-SPFTR Methodologies . . . . .	129
8.2.1	Singular Cases . . . . .	129
8.2.2	Quality of Routing Paths . . . . .	130
8.2.3	Network Throughput . . . . .	131
8.2.4	Computational Cost and Resources Needed . . . . .	132
8.2.5	Summary . . . . .	132
8.3	RFTR Methodology . . . . .	133
8.3.1	Singular Cases . . . . .	134

8.3.2	Quality of Routing Paths . . . . .	134
8.3.3	Resources Required . . . . .	136
8.3.4	Computational Cost . . . . .	136
8.3.5	Network Throughput . . . . .	137
8.3.6	Summary . . . . .	142
8.4	EBR Methodology . . . . .	143
8.4.1	Traffic Patterns and Topologies . . . . .	143
8.4.2	Dropped Packets . . . . .	146
8.4.3	Average Packet Latency . . . . .	146
8.4.4	Network Throughput . . . . .	147
8.4.5	Scalability . . . . .	148
8.4.6	Summary . . . . .	150
<b>9</b>	<b>Conclusions</b>	<b>151</b>
9.1	Conclusions . . . . .	151
9.2	Contributions . . . . .	152
9.3	Publications . . . . .	156
9.4	Future Work . . . . .	157
<b>A</b>	<b>Route Patterns for SPFTR</b>	<b>159</b>
<b>B</b>	<b>Acronyms</b>	<b>167</b>
	<b>Bibliography</b>	<b>169</b>
	<b>Index</b>	<b>183</b>