

# Time Warp from Cluster to Grid

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor  
aan de Universiteit van Amsterdam  
op gezag van de Rector Magnificus  
prof. mr. P. F. van der Heijden  
ten overstaan van een door het college voor promoties ingestelde  
commissie, in het openbaar te verdedigen in de Aula der Universiteit  
op woensdag 8 juni 2005, te 10:00 uur

door

Kamil Antoni Iskra

geboren te Krakau, Polen

---

# Contents

---

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Modelling	1
1.2	Simulation	2
1.3	Parallel discrete event simulation	4
1.3.1	Parallelisation	4
1.3.2	Conservative approach	6
1.3.3	Optimistic approach	7
1.4	Modern computing platforms	9
1.4.1	Cluster computing	9
1.4.2	Grid computing	11
1.5	Objectives	12
<b>2</b>	<b>APSYS Simulation Kernel</b>	<b>15</b>
2.1	Introduction	15
2.2	State of the art	16
2.3	Kernel architecture	19
2.3.1	Interfaces	20
2.3.2	Internal data structures	21
2.4	Simulation models	23
2.4.1	Ising spin	23
2.4.2	PHOLD	25
2.5	Distributed ASCI Supercomputer 2	27

2.6	Process topology	28
2.7	Execution determinism	29
2.7.1	Random number generation	30
2.7.2	Deterministic event ordering	31
2.7.3	Influence on run-time behaviour	32
2.8	Conclusion	34
<b>3</b>	<b>Modelling a Parallel Simulation Kernel</b>	<b>35</b>
3.1	Introduction	35
3.2	State of the art	36
3.3	Self-organised critical behaviour	39
3.3.1	Critical behaviour revisited	41
3.4	Random number generators and optimism	42
3.5	Analytical rollback modelling	45
3.5.1	Rollback length expressed in simulation time	45
3.5.2	Rollback length expressed in events	47
3.5.3	Comparison of the two approaches	49
3.6	Simulator of a parallel kernel	51
3.6.1	Validation	53
3.6.2	Simulating rollbacks	54
3.6.3	Rollback cost	56
3.7	Conclusion	59
<b>4</b>	<b>Event Cancellation Optimisations</b>	<b>61</b>
4.1	Introduction	61
4.2	Cancellation strategies	63
4.2.1	Aggressive cancellation	63
4.2.2	Lazy cancellation	64
4.2.3	Bulk anti-messages	66
4.3	Experiments	68
4.3.1	Single cluster	68
4.3.2	Grid	70
4.4	Conclusion	74
<b>5</b>	<b>Grid-Aware Topology</b>	<b>77</b>
5.1	Introduction	77
5.2	State of the art	78
5.3	Routing processes	82
5.3.1	Process topology	82
5.3.2	Experiments	84

5.3.3	Further improvements	87
5.4	Message aggregation	89
5.4.1	Experiments	90
5.4.2	Aggregation timeout	92
5.5	GVT algorithm	96
5.5.1	Motivation	96
5.5.2	Efficiency improvements	97
5.5.3	Distribution	99
5.6	Conclusion	102
<b>6</b>	<b>Polder Metacomputing Environment</b>	<b>105</b>
6.1	Introduction	105
6.2	State of the art	107
6.3	Task migration	109
6.3.1	Migrating sockets	111
6.3.2	Cross-cluster migration	111
6.3.3	Cross-cluster I/O access	113
6.4	Current efforts	114
<b>7</b>	<b>Conclusion</b>	<b>117</b>
7.1	Summary	117
7.2	Discussion	119
7.3	Future work	121
<b>A</b>	<b>History of APSIS</b>	<b>123</b>
	<b>Nederlandse samenvatting</b>	<b>125</b>
	<b>Streszczenie po polsku</b>	<b>129</b>
	<b>Acknowledgements</b>	<b>133</b>
	<b>Publications</b>	<b>137</b>
	<b>Bibliography</b>	<b>139</b>
	<b>Index</b>	<b>155</b>