

Contents

1	Introduction	1
1.1	Scope of this Work	2
1.2	Document Structure	4
2	Foundations	5
2.1	Resource Management	5
2.2	Grid Computing	7
2.2.1	First Generation	8
2.2.2	Second Generation	8
2.2.3	Third Generation	11
2.2.4	Commercial Grids	13
2.3	Service Level Agreements	13
2.3.1	Basics	15
2.3.1.1	Content of an Agreement	16
2.3.1.2	Initiation of Negotiation	17
2.3.1.3	WS-Agreement and WS-AgreementNegotiation Protocol	17
2.3.2	Structure of a Service Level Agreement	18
2.3.2.1	Context	18
2.3.2.2	Terms	19
2.3.2.3	Guarantee Terms	21
2.3.3	Negotiation of Service Level Agreements	23
3	SLA-aware Scheduling	25
3.1	Negotiation Requirements of Broker Services	25
3.2	Levels of Service Quality	27
3.2.1	SDTs on Resource Specific QoS	28
3.2.2	SDTs on Guarantee Level	29
3.2.3	SDTs on System Policies	31
3.3	Phases of Operation	32
3.3.1	Negotiation	32
3.3.2	Pre-Runtime Phase	34
3.3.3	Runtime Phase	34
3.3.4	Post-Runtime Phase	35

3.4	Timing Aspects of Runtime Phases	36
3.4.1	Overhead caused by Initialization	36
3.4.2	Overhead caused by Checkpointing	38
3.4.3	Overhead caused by Migration	40
3.4.4	Overhead caused by Restart	41
3.4.5	Determining the Minimum Slot	42
3.4.6	Determining the Checkpoint Frequency	42
3.5	Fault Tolerance with intra-cluster scope	46
3.5.1	Basic SLA-aware Scheduling	48
3.5.1.1	Input Parameters	48
3.5.1.2	Output Parameters	49
3.5.1.3	Initialization	51
3.5.1.4	Handling of Running and SLA-Reservation Jobs	52
3.5.1.5	Handling of Problematic Jobs	53
3.5.1.6	Handling of SLA-Requests on Reservations	54
3.5.1.7	Handling of SLA-Non-Reservation Jobs	55
3.5.1.8	Handling of regular SLA-Requests	55
3.5.1.9	Handling of existing and requested Best-Effort Jobs	56
3.5.2	SLA-aware Scheduling with Job Suspension	56
3.5.2.1	Suspension of Best-Effort Jobs	57
3.5.2.2	Suspension of SLA-bound Jobs	61
3.5.3	Partial Execution	62
3.5.3.1	Increasing Level of Fault Tolerance	63
3.5.3.2	Increasing Utilization	64
3.5.4	Buffer Nodes	64
3.6	Fault Tolerance with cross-border scope	65
3.6.1	Identification of Migration Candidates	67
3.6.2	Application of Customer Defined Policies	68
3.6.3	Generation of Compatibility Profiles	69
3.6.4	Static Profile Matching	69
3.6.5	Filtering of Static Results	70
3.6.6	Negotiation and Migration	70
3.6.7	Migration of Problematic Jobs	70
3.6.8	Migration of General Jobs	71

4 Design Aspects 73

4.1	The HPC4U Project	74
4.2	Resource Management System	76
4.2.1	Grid Interface (GI) with Negotiation Manager (NM)	77
4.2.2	Planning Manager (PM) and Machine Manager (MM)	78
4.2.3	Access Manager (AM)	80

4.2.4	Migration Controller (MC)	80
4.2.5	Configuration Manager (CM)	81
4.2.6	Execution Manager (EM)	82
4.2.7	Subsystem Controller (SSC)	82
4.3	Checkpoint Generation	84
4.4	Compatibility Profile	87
4.4.1	Architecture and System Properties	88
4.4.1.1	Operating System	88
4.4.1.2	Checkpoint System	89
4.4.1.3	Processor Architecture	89
4.4.1.4	Other System Properties	90
4.4.2	Libraries	90
5	Results and Perspectives	93
5.1	Fault Tolerance Provision	94
5.2	Use-Case Experiences	98
5.2.1	Swedish Meteorological and Hydrological Institute	98
5.2.1.1	Fault Tolerance Aspects	100
5.2.2	Centre of Excellence in Aeronautical Research	101
5.2.2.1	Fault Tolerance Aspects	101
5.3	Key Requirement Security	101
6	Related Work	105
6.1	Service Level Agreements in Grid Middleware	105
6.1.1	Advance Reservations and GARA	105
6.1.2	Service Negotiation and Allocation Protocol	107
6.1.3	Grid Broker Services	109
6.1.3.1	GRUBER/DI-GRUBER SLA broker	109
6.1.3.2	White Rose Grid	110
6.1.3.3	NextGRID	111
6.1.3.4	SLBSH Project	112
6.1.4	Grid Economy	113
6.2	Process Checkpointing	114
6.2.1	Berkeley Labs Checkpointing and Restart (BLCR)	115
6.2.2	IBM Metacluster	116
7	Conclusion	119
	Bibliography	123