

Contents

List of Figures	iv
1 Introduction	1
1.1 Scheduling and Performance Issues	3
1.2 Motivation for our work	4
1.3 Goals of our work	5
1.4 Organization of the Thesis	6
2 Background	9
2.1 Cloud Computing Paradigm	9
2.2 Cloud Computing Architecture	11
2.2.1 Essential Characteristics	11
2.2.2 Components of Cloud Computing	12
2.2.3 Layers of Cloud Model	13
2.2.4 Cloud Service Models	14
2.2.5 Deployment Models	16
2.3 Cloud Resource Virtualization	19
2.3.1 Virtualization	19
2.3.2 Virtual Machine	21
2.3.3 Virtual Machine Monitors	22
2.3.4 Cloud Provisioning Approach	22
2.3.5 Virtual Machines Provisioning & Manageability	23
2.4 Performance Management on Cloud	27
2.4.1 Information Sharing	27

2.4.2	Heterogeneous Environment	28
2.4.3	Unpredictability	28
2.5	Related Work	29
3	Dynamic Allocation of Virtual Machines in Cloud Center	31
3.1	Introduction	31
3.2	System Model	32
3.3	Analysis of the Model	35
3.4	Numerical Illustration	39
3.5	Conclusion	46
4	QoS of Private Cloud in Finite Population Environment	47
4.1	Introduction	47
4.2	Service Policy 1	49
4.2.1	Performance indices	51
4.2.2	Numerical Illustration	52
4.3	Service Policy 2	56
4.3.1	Performance indices	58
4.3.2	Numerical Illustration	60
4.4	Conclusion	65
5	QoS of Cloud Centers With Different Arrival Modes	67
5.1	Introduction	67
5.2	System Description	68
5.3	Model description and its analysis	71
5.4	System performance measures	73
5.4.1	Cost analysis	74
5.5	Numerical Illustration	76
5.6	Conclusion	83
6	Dynamic Resource Provisioning in Multi-tier Applications	85
6.1	Introduction	85
6.2	Architecture Overview	86

<i>CONTENTS</i>	iii
6.2.1 Cloud Computing Infrastructure	87
6.2.2 Virtualized Multi-tier Application Queueing Model	89
6.2.3 Active Monitoring Load Balancer	89
6.3 Model Description and Analysis	90
6.3.1 Relation between steady-state distribution at arbitrary and pre-arrival epochs	91
6.4 Computational algorithm	92
6.5 Performance Measures	94
6.6 Numerical Illustrations	94
6.7 Conclusion	100
7 Bulk Services in Cloud Computing Centers	101
7.1 Introduction	101
7.2 System Model	102
7.3 Modelling and Analysis	105
7.4 Performance measures	107
7.4.1 Waiting time analysis	108
7.5 Numerical Illustrations	109
7.6 Conclusion	115
8 Conclusions	117
List of Publications from the Thesis	120
BIBLIOGRAPHY	123