

Economic Models for Managing Cloud Services

Sajib Mistry • Athman Bouguettaya • Hai Dong

Economic Models for Managing Cloud Services



Springer

Sajib Mistry
School of Information Technologies
University of Sydney
Sydney, NSW, Australia

Athman Bouguettaya
School of Information Technologies
University of Sydney
Sydney, NSW, Australia

Hai Dong
School of Science
RMIT University
Melbourne, VIC, Australia

ISBN 978-3-319-73875-8 ISBN 978-3-319-73876-5 (eBook)
<https://doi.org/10.1007/978-3-319-73876-5>

Library of Congress Control Number: 2017964374

© Springer International Publishing AG 2018

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Printed on acid-free paper

This Springer imprint is published by the registered company Springer International Publishing AG part of Springer Nature

The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

*To my parents, Sunil and Mili, and my wife,
Pragga, and my daughter, Aapti.*

Sajib Mistry

*To my wife and best companion and friend,
Malika.*

Athman Bouguettaya

*To my parents, Jianwen Dong and Aiying
Yang.*

Hai Dong

Foreword

Cloud computing is inexorably making large inroads to becoming the technology of choice among small and large businesses via provisioning IT infrastructure resources and hosting applications. A significant advantage of cloud computing is its economic benefits to both service consumers and providers. From a consumer's perspective, cloud computing is a model for providing computing resources as a location-independent and highly scalable service that is acquired on demand with little or no fixed capital investment. Cloud is ideal for those organizations with fluctuating computational resource demands. It is able to fulfill those demands with dynamically provisioned resources. From a provider's perspective, cloud computing provides economies of scale through a distribution of costs among a large pool of consumers, centralization of infrastructures in locations with lower costs, and improved resource utilization.

Cloud service management is a critical aspect of cloud computing. In cloud computing, a complex business process can be realized by outsourcing its involved tasks to the cloud and then composing the resulting component cloud services against each of these outsourced tasks. The cloud service management from the consumers' perspective aims to select the optimal provider for each component cloud service to minimize consumers' total cost while meeting consumers' Quality of Service (QoS) requirements. Similarly, cloud providers require an effective cloud service management process to achieve their business goals, e.g., specific revenue or profit expectations. In this regard, the cloud service management process can be realized by selecting and composing more profitable services to fulfill consumers' demands while ensuring delivery of the promised QoS. From here, we can see that cloud computing management is mainly driven by economic factors. Employing economic models to design the cloud service management framework is a right way, as it perfectly fits the way in which business is conducted.

There are a variety of books on the market that cover many interesting issues related to cloud computing. However, none but this book provide a comprehensive analysis of economic models in the long-term cloud service management from the providers' perspective. This book is a first attempt to design quantitative and qualitative economic models for an Infrastructure as a Service (IaaS) provider. It provides

an efficient, market-driven, and competitive environment for cloud providers and consumers to maximize their profit and minimize their costs, respectively, while meeting their long-term QoS requests. This is crucial if we want to unlock the full potential of cloud computing which has largely remained closed and proprietary. This book also covers fundamental technical details concerning optimization, prediction, and machine learning models in the cloud service management. The state-of-the-art technologies described and the references included in this book will also help the interested readers gain knowledge on these topics.

This book can be utilized as a useful reference to anyone who is interested in theory, practice, and application of economic models in cloud computing. This book will be an invaluable guide for small and medium entrepreneurs who have invested or plan to invest in cloud infrastructures and services. Overall, this book is suitable for a wide audience that includes students, researchers, and practitioners studying or working in service-oriented computing and cloud computing.

School of Computer Science and Engineering
The University of New South Wales
Sydney, NSW, Australia
September 18, 2017

Professor Boualem Benatallah

Preface

Cloud computing is increasingly becoming the technology of choice as the next-generation platform for conducting businesses. Cloud solutions are represented as services, i.e., a higher level abstraction of computing resources. An effective cloud service management framework has the potential of creating a sustainable cloud service market. It enables the wider adoption of the cloud at a greater scale and faster pace by considering economic perspectives of service consumers and providers. The long-term cloud service composition is an essential element to designing an effective management framework. The composition from the consumer's perspective aims to minimize total costs while meeting Quality of Service (QoS) requirements by selecting the best set of cloud providers. The composition from the provider's perspective aims to maximize profit for a long-term period by selecting an optimal set of service requests. Developing an efficient and long-term cloud service composition framework from the provider's perspective is very important to sustaining and growing demand from a large number of service consumers in the market.

In this book, we develop an economic model-driven long-term cloud service composition framework from the Infrastructure as a Service (IaaS) provider's perspective. The provider-consumer relationships in a cloud market are usually long term and economically driven. In this regard, using economic models to optimize the composition of service requests fits perfectly well with the way in which business is usually conducted. First, a new quantitative economic model is developed that maximizes the provider's long-term revenue and profit by selecting an optimal set of IaaS requests in a dynamic economic environment. We propose a new multivariate Hidden Markov and Autoregressive Integrated Moving Average (HMM-ARIMA) model to predict various patterns of runtime resource utilization. A heuristic-based Integer Linear Programming (ILP) optimization approach is proposed to maximize the runtime resource utilization. We deploy a Dynamic Bayesian Network (DBN) to model the dynamic pricing and long-term operation cost. A new Hybrid Adaptive Genetic Algorithm (HAGA) is developed that optimizes a nonlinear profit function periodically to address the stochastic arrival of requests. Next, we develop a qualitative economic model which is a preference-driven approach to enact on high-

level business strategies, such as service menu creation and pricing. We propose the Temporal Conditional Preference Network (TempCP-Net) to represent the high-level IaaS business strategies. The temporal qualitative preferences are indexed in a multidimensional k -d tree to efficiently compute the preference ranking in runtime. A three-dimensional Q-learning approach is proposed to find an optimal qualitative composition using statistical analysis on historical request patterns. Finally, we propose a new multivariate approach to predict future Quality of Service (QoS) performances of peer service providers to efficiently configure a TempCP-Net.

We have evaluated the efficiency of the proposed framework using Google Cluster data, real-world QoS data, and synthetic data. Experimental results show that the proposed composition framework efficiently maximizes the provider's long-term economic goals in runtime. The proposed models in this book are expected to play a significant role in creating an economically viable and stable cloud market.

Sydney, NSW, Australia
Sydney, NSW, Australia
Melbourne, VIC, Australia

Sajib Mistry
Athman Bouguettaya
Hai Dong

Acknowledgments

I owe a huge debt of gratitude to my wife, without whose love and support, I would not have finished this book. I would also like to thank my parents, parents-in-law, brother, and sister-in-law for their encouragement. I would also like to thank all my friends who directly or indirectly supported me.

Sydney, NSW, Australia

Sajib Mistry

I would like to thank my family for their unwavering support during my work on this book.

Sydney, NSW, Australia

Athman Bouguettaya

I would like to acknowledge with gratitude the love and support of my family members and friends along the way.

Melbourne, VIC, Australia

Hai Dong

The authors of this book would like to extend their sincere gratitude and appreciation to their collaborators for the contribution to this book. In particular, we would like to acknowledge Dr. Kai Qin, Dr. Abdelkarim Erradi, and other collaborators in the Sensor Cloud Services Laboratory (SCSLab) at the University of Sydney.

Contents

- 1 Introduction 1**
 - 1.1 Cloud Service Management 2
 - 1.1.1 Cloud Management Using Service Composition..... 5
 - 1.2 Economic Models for Better Cloud Service Management 7
 - 1.2.1 Challenges in Developing a Quantitative Economic Model .. 8
 - 1.2.2 Challenges in Developing a Qualitative Economic Model ... 10
 - 1.2.3 Economic Model Based Cloud Service Composition..... 12
 - 1.3 Outline of the Book Chapters..... 14
- 2 Background 17**
 - 2.1 Cloud Service Management from an End User’s Perspective..... 17
 - 2.1.1 Service Composition with Functional Requirements 17
 - 2.1.2 Service Composition with Non-functional Requirements 18
 - 2.1.3 Service Composition with Long-Term Requirements..... 19
 - 2.2 Cloud Service Management from a Provider’s Perspective 20
 - 2.2.1 Resource Allocation Approaches 20
 - 2.2.2 Task Scheduling Approaches..... 21
 - 2.2.3 Admission Control Approaches..... 22
 - 2.3 Economic Models 22
 - 2.3.1 Economic Modeling in Operations Research..... 23
 - 2.3.2 Quantitative Economic Modeling in the Cloud Market 24
 - 2.3.3 Qualitative Economic Modeling in the Cloud Market 25
 - 2.4 Prediction Modeling in Service Composition 26
 - 2.4.1 Time-Series and Probabilistic Prediction Models 26
 - 2.4.2 Web Service QoS Prediction Frameworks..... 27
 - 2.5 Optimization Approaches in Service Composition 28
 - 2.5.1 Global Optimization Approaches 28
 - 2.5.2 Sequential Local Optimization and Machine-Learning Approaches..... 30
 - 2.6 Conclusion 31

3	Long-Term IaaS Composition for Deterministic Requests	33
3.1	Introduction	33
3.2	The Heuristics on Consumer Behavior	36
3.3	The Long-Term Composition Framework for Deterministic Requests	37
3.4	Predicting the Dynamic Behavior of Consumer Requests	38
3.4.1	Predicting Runtime Behavior of Existing Consumers' Requests	39
3.4.2	Predicting Runtime Behavior of New Consumers' Requests	44
3.5	An ILP Modeling for Request Optimization	45
3.6	Experiments and Results	47
3.6.1	Data Description	47
3.6.2	Accuracy in Predicting the Behavior of Consumer Requests	49
3.6.3	Performance Analysis on Profit Maximization	51
3.7	Conclusion	51
4	Long-Term IaaS Composition for Stochastic Requests	53
4.1	Introduction	53
4.2	Long-Term Dynamic IaaS Composition Framework	55
4.3	Long-Term Economic Model of IaaS Provider	58
4.3.1	Long-Term Economic Valuation	58
4.3.2	Semantic Economic Expectation and Fitness of a Composition	61
4.4	Genetic Optimization Using IaaS Economic Model	63
4.5	Hybrid Adaptive Genetic Algorithm (HAGA) Based Composition	64
4.5.1	Solution Representation in HAGA	65
4.5.2	Initial Population Generation	66
4.5.3	Parent Selection, Crossover, and Mutation	66
4.5.4	Solution Generation with Repair Heuristic	66
4.5.5	Runtime Optimization Scheduling	68
4.6	Experiments and Results	70
4.6.1	Data Description	71
4.6.2	Setup of Long-Term Economic Model	71
4.6.3	Setup of HAGA, GA, and ACO Parameters	72
4.6.4	Efficiency of the HAGA-Based Composition	72
4.6.5	Time Complexity Analysis	75
4.7	Conclusion	76
5	Long-Term Qualitative IaaS Composition	77
5.1	Introduction	77
5.2	Motivation: A Qualitative IaaS Economic Model with Decision Variables	79
5.3	The Temporal CP-Net Based Qualitative Economic Model	83
5.3.1	The k -d Tree Indexing of the Induced Preference Graph	85
5.3.2	Ranking the Consumers' Requests Using the TempCP-Net	86

5.4	Optimization Algorithms for Qualitative IaaS Composition	87
5.4.1	Dynamic Programming-Based IaaS Composition	88
5.4.2	Heuristic-Based Sequential IaaS Composition	89
5.5	Reinforcement Learning for Long-Term IaaS Requests Composition	90
5.5.1	A MDP Model for Long-Term Composition	91
5.5.2	Q-learning Approach for IaaS Composition	93
5.5.3	Determining Initial Q-values Using Statistical Analysis.....	97
5.6	Experiments and Results.....	99
5.6.1	Simulation Setup.....	100
5.6.2	Efficiency of the Heuristic-Based Sequential Optimization ..	101
5.6.3	Time Complexity Analysis of the Sequential Optimization ..	103
5.6.4	Efficiency of the Q-learning Based IaaS Composition.....	104
5.6.5	Runtime Efficiency of the Heuristic-Based Learned Q-matrices.....	107
5.7	Conclusion	110
6	Service Providers' Long-Term QoS Prediction Model.....	111
6.1	Introduction	111
6.2	The Multivariate QoS Forecasting Framework	112
6.3	Multivariate QoS Prediction Model (MQPM)	113
6.3.1	Transforming ARIMA into the MQPM.....	115
6.3.2	Transforming Holt-Winters into the MQPM	115
6.3.3	Prediction Error Reduction Algorithm in the MQPM.....	116
6.4	Forecasting from the MQPM	117
6.5	Experiments and Results.....	119
6.5.1	Data Description	119
6.5.2	Comparison Among the Prediction Models	119
6.5.3	Effect of the Time-Series Length on MQPM.....	121
6.6	Conclusion	122
7	Conclusion	123
7.1	Future Work	130
	References.....	133

List of Figures

Fig. 1.1	Long-term consumer requirements for (a) three resource attributes, and (b) three QoS attributes	4
Fig. 1.2	Arrival models of incoming requests. (a) One-time arrival of requests. (b) Deterministic arrival of requests. (c) Stochastic arrival of requests	5
Fig. 1.3	A cloud service composition architecture	6
Fig. 3.1	The long-term resource utilization patterns: (a) High-frequent, and (b) Seasonal-trending	35
Fig. 3.2	The proposed service composition framework	38
Fig. 3.3	(a) Original and transformed QoS requirements, and (b) Resource requirements	39
Fig. 3.4	Correlation among QoS attributes in the multivariate HMM	41
Fig. 3.5	Prediction accuracy on Google Cluster Data	49
Fig. 3.6	CDI effect on the proposed approach	50
Fig. 3.7	Prediction accuracy of the community heuristic	50
Fig. 3.8	Effect of the community size on prediction	51
Fig. 3.9	Cumulative profit over time	52
Fig. 3.10	Monthly resource utilization	52
Fig. 4.1	The proposed dynamic service composition framework	55
Fig. 4.2	Original and transformed long-term resource requirements	56
Fig. 4.3	A DBN of dynamic pricing	59
Fig. 4.4	A weighted long-term economic expectation	63
Fig. 4.5	Hybrid genetic optimization in IaaS composition	65
Fig. 4.6	Runtime modification of the inter-dependency heuristic	70
Fig. 4.7	Prediction accuracy of DBN	73
Fig. 4.8	Effect of request arrival intensity in solution convergence	73
Fig. 4.9	Comparison of HAGA-based long-term composition with other approaches	74
Fig. 4.10	Effect of the solution space size	75

Fig. 4.11	Number of iterations in dynamic environment	75
Fig. 4.12	Global solution convergence time	76
Fig. 5.1	(a) Semantic representation of preferred service attributes, and (b) a TempCP-Net.....	80
Fig. 5.2	(a) Incoming requests, and (b) the preference ranking table.....	82
Fig. 5.3	The induced preference graph of CP1	84
Fig. 5.4	The k -d tree indexing of the induced preference graph	85
Fig. 5.5	(a) Temporal semantic segmentation of a request, and (b) some sequential orders for local optimization.....	87
Fig. 5.6	The key composition scenarios (a) almost disjoint, (b) almost overlapping, (c) chain, and (d) hybrid patterns	89
Fig. 5.7	A MDP model for IaaS request composition	92
Fig. 5.8	(a) 2d Q-values in a $Q[S,A]$, and (b) 3d Q-values in a $Q[S,A, O]$...	96
Fig. 5.9	Different types of distributions: (a) Normal, (b) right-skewed, and (c) left-skewed	98
Fig. 5.10	Accuracy of the heuristic-based sequential approach in the disjoint pattern	102
Fig. 5.11	Accuracy of the heuristic-based sequential approach in the overlapping pattern	102
Fig. 5.12	Accuracy of the heuristic-based sequential approach in the chain pattern	103
Fig. 5.13	Accuracy of the heuristic-based sequential approach in the hybrid pattern	103
Fig. 5.14	The global vs sequential optimization time complexity	104
Fig. 5.15	Accuracy of the Q-learning approach in the normal distribution	105
Fig. 5.16	Accuracy of the Q-learning approach in the right-skewed distribution	105
Fig. 5.17	Accuracy of the Q-learning approach in the left-skewed distribution	106
Fig. 5.18	Accuracy of the Q-learning approach in the random distribution	106
Fig. 5.19	Runtime accuracy of the proposed approach for similar pattern	107
Fig. 5.20	Runtime accuracy of the proposed approach for dissimilar pattern	108
Fig. 5.21	Runtime performance in different distributions	109
Fig. 5.22	Time complexity in different distributions	109
Fig. 6.1	The multivariate QoS prediction framework	113
Fig. 6.2	Prediction error reduction using a multivariate analysis	114
Fig. 6.3	RMSE of throughput in different service providers	120
Fig. 6.4	RMSE of response time in different service providers	120
Fig. 6.5	Effect of time-series length on RMSE	122
Fig. 7.1	Key elements in the long-term IaaS composition	124

List of Tables

Table 2.1	Different prediction models in the service composition	27
Table 2.2	Different global optimization approaches in the service composition	29
Table 3.1	Summary of Google Cluster Dataset	47
Table 3.2	Relationship between resource utilization and operation cost.....	48
Table 4.1	Requested and initially transformed resource usage as well as runtime update	57
Table 4.2	Economic benefits in composition plans	57
Table 4.3	Distance function for semantic expectations	62
Table 4.4	Long-term relationship between utility factor and operation cost..	72
Table 4.5	HAGA, GA, and ACO parameter values	72
Table 5.1	Four distributions according to the request length	100
Table 6.1	Summary of the real-world QoS dataset	119
Table 6.2	Summary of the performance of prediction models	121