

POLITECNICO DI MILANO

Scuola di Ingegneria dell'Informazione
Corso di Laurea Magistrale in Ingegneria Informatica
Dipartimento di Elettronica, Informazione e Bioingegneria



**Generalized Nash Equilibria
for the Service Provisioning Problem
in Multi-Cloud Systems**

Advisor: Prof. Danilo ARDAGNA

Co-Advisor: Prof. Mauro PASSACANTANDO

Master Thesis by:
Ettore TREVISIOL - 770628

Academic Year 2012/2013

Contents

1	Introduction	5
2	State of the art	9
2.1	Cloud Computing basic concepts	9
2.2	Cloud Computing definition	11
2.2.1	Characteristics	13
2.2.2	Structure models	14
2.3	Cloud Computing and run-time research challenges	23
2.3.1	Problem	25
2.3.2	Solution	25
2.3.3	Discipline	26
2.3.4	State of the art	27
2.3.5	Classification of the state of the art	33
2.3.6	Criteria for evaluation	38
3	A game theory service provisioning model	41
3.1	Problem statement and assumptions	41
3.2	Generalized Nash game model	44
3.3	Game analysis	48
3.3.1	Dominant strategies for IaaS	48
3.3.2	Game potential	48
3.3.3	Analysis of constraints	49
3.3.4	Game model reformulation	52
3.4	A distributed algorithm for identifying Generalized Nash Equilibria	53
4	Tools	59
4.1	AMPL	59
4.2	CPLEX	60
4.2.1	CPLEX algorithms for continuous optimization	60
4.3	SPECweb 2005	61

CONTENTS

4.4	JMeter	63
4.5	SPECweb deployment in the Cloud	65
4.5.1	SPECweb tests	65
4.5.2	JMeter extension	67
4.5.3	SPECmeter	69
4.6	Cloud analysis tool	71
4.6.1	Cloud analysis tool class diagram	71
4.6.2	Cloud analysis tool sequence diagram	72
5	Experimental results	75
5.1	Design of experiments	75
5.1.1	Parameters generation	75
5.1.2	SaaS to IaaS mapping	77
5.1.3	Traffic generation	77
5.2	Scalability analysis	80
5.3	Equilibria efficiency	81
5.3.1	Social optimum problem	82
5.3.2	Price of Anarchy and Individual Worst Case	83
5.4	Alternative algorithms for resource allocation	83
5.4.1	Heuristic	84
5.4.2	Resource rescaling algorithm	89
5.5	Algorithms efficiency comparison	90
5.6	Multiple IaaS analysis	96
6	Conclusions	105
	Appendices	109
A	Game theory and generalized Nash equilibrium problem	109
A.1	Game theory in the Cloud Computing	109
A.2	Definition of Game	110
A.3	Solution concepts: Nash Equilibrium and Generalized Nash Equilibrium	111
A.4	Equilibria existence and potential games	114
A.5	Wardrop equilibrium	117

List of Figures

2.1	Cloud Computing architecture, [106].	15
2.2	Cloud service models.	18
2.3	Cloud deployment models.	22
2.4	Taxonomy for optimization approaches.	28
3.1	Cloud infrastructures.	42
3.2	System performance model.	43
3.3	Algorithm for finding Generalized Nash Equilibria.	55
4.1	SPECweb 2005 test diagram.	66
4.2	SPECweb banking test Markov chain.	67
4.3	Markov chain example in JMeter.	68
4.4	SPECmeter architecture.	69
4.5	SPECmeter test automation sequence diagram.	70
4.6	Cloud analysis tool class diagram.	72
4.7	Cloud analysis tool sequence diagram.	73
5.1	Queueing delay time.	76
5.2	Service time.	76
5.3	Daily time distribution of requests.	78
5.4	Weekly time distribution of requests.	79
5.5	Generated daily time distribution of requests.	79
5.6	Worldwide distribution of requests.	80
5.7	Distributed algorithm for identifying GNE scalability.	82
5.8	Impact of system capacity and service classes on the optimal profit, [59].	84
5.9	The difference of optimal profit with varied number of opened service classes, [59].	85
5.10	IaaS capacity usage on peak hours.	92
5.11	Traffic example of a single SaaS provider.	97
5.12	SaaS allocation with unlimited resources.	98

LIST OF FIGURES

5.13 SaaS allocation with limited resources.	99
5.14 Multi-IaaS analysis results with $\phi_i = 0.1$	100
5.15 Multi-IaaS payoff function comparison with $\phi_i = 0.1$	101
5.16 Multi-IaaS analysis results with $\phi_i = 0.3$	102
5.17 Multi-IaaS payoff function comparison with $\phi_i = 0.3$	103
5.18 Multi-IaaS analysis results with $\phi_i = 0.5$	103
5.19 Multi-IaaS payoff function comparison with $\phi_i = 0.5$	104
A.1 Families of Generalized Nash Equilibrium Problems.	114

List of Tables

2.1	Problem category: perspective.	33
2.2	Problem category: quality attributes.	34
2.3	Problem category: dimensionality.	34
2.4	Problem category: constraints.	35
2.5	Solution category: type.	35
2.6	Solution category: degrees of freedom.	36
2.7	Solution category: architecture representation.	36
2.8	Solution category: optimization strategy.	36
2.9	Solution category: constraints handling.	37
2.10	Solution category: time scale.	37
2.11	Discipline category: type.	37
2.12	Discipline category: quality model.	37
3.1	Parameters and decision variables.	47
5.1	Performance parameters and time unit costs.	77
5.2	Distributed algorithm for identifying GNE execution times.	81
5.3	Algorithms PoA comparison with $\phi_i = 0.1$	93
5.4	Algorithms IWC comparison with $\phi_i = 0.1$	93
5.5	Algorithms re-optimization comparison with $\phi_i = 0.1$	93
5.6	Algorithms PoA comparison with $\phi_i = 0.3$	94
5.7	Algorithms IWC comparison with $\phi_i = 0.3$	94
5.8	Algorithms re-optimization comparison with $\phi_i = 0.3$	94
5.9	Algorithms PoA comparison with $\phi_i = 0.5$	95
5.10	Algorithms IWC comparison with $\phi_i = 0.5$	95
5.11	Algorithms re-optimization comparison with $\phi_i = 0.5$	95
5.12	Multi-IaaS analysis results with $\phi_i = 0.1$	99
5.13	Multi-IaaS analysis results with $\phi_i = 0.3$	100
5.14	Multi-IaaS analysis results with $\phi_i = 0.5$	101

LIST OF TABLES

Abstract

In recent years the evolution and the widespread adoption of virtualization, service-oriented architectures, autonomic, and utility computing have converged letting a new paradigm to emerge: Cloud Computing. Clouds allow the on-demand delivering of software, hardware, and data as services. Currently the Cloud offer is becoming day by day wider, since all the major IT companies and Service providers, like Microsoft, Google, Amazon, HP, IBM, and VMware have started providing solutions involving this new technological paradigm.

As Cloud-based services are more numerous and dynamic, the development of efficient service provisioning policies becomes increasingly challenging. In this thesis we take the perspective of Software as a Service (SaaS) providers which host their applications at multiple Infrastructure as a Service (IaaS) providers. Each SaaS needs to comply with quality of service requirements, specified in Service Level Agreement (SLA) contracts with the end-users, which determine the revenues and penalties on the basis of the achieved performance level. Each SaaS provider wants to minimize the cost of use of Cloud resources and penalties for requests execution failures. Moreover, SaaS providers compete and bid for the use of infrastructural resources. On the other hand, the IaaS want to maximize their revenues obtained providing virtualized resources.

In this thesis we model the service provisioning problem as a generalized Nash game. In particular, we develop an efficient distributed algorithm for the run-time allocation of IaaS resources among competing SaaS providers. We demonstrate the effectiveness of our approach by performing tests considering realistic Cloud scenarios. Numerical results show that our algorithm is scalable and can be used at run-time since it can solve problem instances of maximum size in less than one minute. Compared to other state-of-the-art solutions our model can improve the efficiency of Cloud system evaluated in term of Price of Anarchy up to 100%. Furthermore our analyses point out the SaaS benefits while exploiting multiple IaaS deployment of applications and redistribution of application workloads.

Sommario

Negli ultimi anni l'evoluzione e la diffusa adozione di virtualizzazione, architetture orientate ai servizi, autonomic e utility computing sono confluiti in un nuovo paradigma: il Cloud Computing. Il Cloud Computing ha come obiettivo la fornitura on-demand di software e hardware come risorse accessibili tramite Internet.

Con l'aumento della quantità e della dinamicità dei servizi basati sul Cloud, lo sviluppo di politiche efficienti per la distribuzione delle risorse è diventato sempre più complesso. In questo lavoro di tesi abbiamo studiato il problema dal punto di vista dei fornitori Software as a Service (SaaS) che ospitano le loro applicazioni presso molteplici fornitori Infrastructure as a Service (IaaS). Ogni SaaS deve rispettare la qualità del servizio, specificata nei contratti di Service Level Agreement (SLA) con i propri clienti, che determina i ricavi e le penalità sulla base del livello di prestazioni raggiunto. Ogni SaaS vuole minimizzare i costi di utilizzo delle risorse Cloud e delle sanzioni causate dalla violazione dei contratti di SLA. Inoltre, i fornitori SaaS competono fra loro facendo offerte per l'utilizzo delle infrastrutture. Al contrario, gli IaaS vogliono massimizzare i propri introiti ottenuti fornendo risorse virtualizzate.

In questa tesi abbiamo modellato il problema per la fornitura delle risorse come un gioco di Nash generalizzato. In particolare, abbiamo sviluppato un algoritmo distribuito per la gestione a run-time delle risorse degli IaaS fra i SaaS in competizione. Abbiamo dimostrato l'efficacia del nostro approccio compiendo test rappresentativi di scenari di carico reale. I risultati numerici mostrano che l'algoritmo è scalabile e può essere utilizzato a run-time, poiché può risolvere istanze del problema di dimensione massima in meno di un minuto. Rispetto ad altre soluzioni della letteratura il nostro modello può migliorare l'efficienza in termini di Price of Anarchy del sistema Cloud valutato fino al 100%. Inoltre, le nostre analisi evidenziano i benefici che i SaaS possono ottenere sfruttando il dislocamento delle applicazioni e la distribuzione del traffico su molteplici IaaS.

Chapter 1

Introduction

Cloud Computing has been a dominant IT news topic over the past few years. It is essentially a way for IT companies to deliver software/hardware on-demand as services through the Internet. Cloud Computing applications are generally priced on a subscription model, so end-users may pay a yearly usage fee, for example, rather than the more familiar model of purchasing software licenses. The Cloud-based services are not only restricted to software applications (Software as a Service – SaaS), but could also be the platform for the deployment and execution of applications developed in house (Platform as a Service – PaaS) and the hardware infrastructure (Infrastructure as a Service – IaaS).

In the SaaS paradigm, applications are available over the Web and provide Quality of Service (QoS) guarantees to end-users. The SaaS provider hosts both the application and the data, hence the end-user is able to use and access the service from all over the world. With PaaS, applications are developed and deployed on platforms transparently managed by the Cloud provider. The platform typically includes databases, middleware, and also development tools. In IaaS systems, virtual computer environments are provided as services and servers, storage, and network equipment can be outsourced by customers without the expertise to operate them.

Many companies, e.g., Google, Amazon, and Microsoft are offering Cloud Computing services such as Google’s App Engine and Amazon’s Elastic Compute Cloud (EC2) or Microsoft Windows Azure. Large data centers provide the infrastructure behind the Cloud and virtualization technology makes Cloud Computing resources more efficient and cost-effective both for providers and customers. Indeed, end-users obtain the benefits of the infrastructure without the need to implement and administer it directly adding or removing capacity almost instantaneously on a “pay-as-you-use” basis. Cloud providers can, on the other hand, maximize the utilization of their

physical resources also obtaining economies of scale.

The development of efficient service provisioning policies is among the major issues in Cloud research. Indeed, modern Clouds operate in a new and dynamic world, characterized by continuous changes in the environment and in the system and performance requirements must be satisfied. Continuous changes occurs without warning and in an unpredictable manner, and are outside the control of the Cloud provider. Therefore, advanced solution need to be developed to manage the Cloud system in a dynamically adaptive way, while continuously providing service and performance guarantees.

The recent evolution of Cloud system and the rapid growth of the Internet have led to a remarkable usage of game-theoretic tools. Problems arising in the ICT industry, such as resource or quality of service allocation problem, pricing, and load shedding, can not always be handled with classical optimization approaches because each player can be affected by the action of all players, not only by his own actions. In this context, game theory models and approaches allow to gain an in-depth analytical understanding of the service provisioning problem.

Game Theory has been successfully applied to diverse problems such as Internet pricing, flow and congestion control, routing, and networking. One of the most widely used “solution concept” in Game Theory is the Nash Equilibrium approach [71]: a set of strategies for the players constitute a Nash Equilibrium if no player can benefit by changing his/her strategy unilaterally or, in other words, every player is playing a *best response* to the strategy choices of his/her opponents.

In this thesis we take the perspective of SaaS providers which host their applications at multiple IaaS providers, thanks to a software layer developed within the MODAClouds project [68] [13]. Each SaaS provider wants to minimize the cost of use of Cloud resources and incurs in penalties in case of requests execution failures. The cost minimization is challenging since online services receive dynamic workloads that fluctuate during the day. Resources have to be allocated flexibly at run-time according to workload fluctuations. Furthermore, each SaaS behaves selfishly and competes with others SaaS for the use of infrastructural resources supplied by the IaaS. Each IaaS, in his turn, wants to maximize the revenues obtained providing the resources.

To capture the behavior of SaaSs and IaaSs in this conflicting situation in which the best choice for one depends on the choices of the others, we recur to the *Generalized Nash Equilibrium* (GNE) concept, which is an extension of the classical Nash equilibrium.

Therefore, the run-time service provisioning problem will be modeled as a *Generalized Nash Equilibrium Problem* (GNEP). We then use Noncooperative Game Theory results to develop an efficient algorithm for the run-time

management and allocation of IaaS resources to competing SaaS suitable also for a *fully distributed* implementation. Multiple solutions achieving generalized equilibria will be proposed and evaluated in terms of their efficiency with respect to the *social optimum* of the Cloud. We will demonstrate the effectiveness of our approach by simulation and performing tests on a real prototype environment.

The remainder of the thesis is organized as follows:

- In Chapter 2 we present a general overview on Cloud Computing providing definitions, illustrating the main characteristics and showing the different structures models available. Afterwards, we will explain the state of the art concepts and techniques relative to our work. An analysis and a classification of the literature approaches is given in terms of type of problem, solution found and discipline adopted, according to the approach used: pure optimization or game theory.
- In Chapter 3 the game-theoretic service provisioning problem will be faced. We will start introducing the problem statements and design assumptions. Then an analysis of the game and its properties will be performed. Finally, we will present an algorithm able to find an equilibrium for the resource allocation problem.
- In Chapter 4 we will describe the tools used in this thesis work. A description of the optimization modelers and solvers we adopted will be provided. Furthermore, we will describe the workload injector we have developed to estimate the performance parameters of Cloud applications.
- Chapter 5 will be dedicated to assess the quality of our solution through analyses and experiments. After a description of experiments settings, algorithm scalability and comparison with other two approaches will be performed. Finally, we will analyze the benefits that can be achieved by SaaS when hosting applications on multiple IaaS.
- In Chapter 6 conclusions will be drawn, underling the achieved results and presenting future research directions.

Chapter 6

Conclusions

As Cloud-based services become more numerous and dynamic, resource provisioning becomes more and more challenging, especially when decisions can be affected by the action of others, not only by own actions, hence requiring game-theory approaches. Indeed, in any time instant resources have to be allocated to handle effectively workload fluctuations, while providing quality of service guarantees to the end users.

The overall goal we addressed in our thesis is the minimization of the costs associated with the allocated virtual machine instances in multiple IaaSs, while guaranteeing QoS constraints. To achieve this purpose we have proposed a game-theoretic approach for the run-time management of IaaSs provider capacities among multiple competing SaaSs. The cost model consists of objective functions which include revenues and penalties incurred depending on the achieved performance level and infrastructural costs associated with IaaSs resources. Therefore a distributed algorithm for identifying Generalized Nash Equilibria have been presented and its termination in a finite number of iterations has been demonstrated.

Thanks the AMPL language and the CPLEX solver, the effectiveness of our approach have been assessed by performing a wide set of analyses under multiple workload conditions. Realistic workloads created from a large website statistics and performance parameters estimated on an industrial benchmarking deployed in the Cloud have been used.

A number of different scenario of interest have been considered. Systems up to thousands of applications can be managed very efficiently in a fully distributed manner. Our algorithm found efficient GNE, for a hourly basis resource allocation, in less than a minute, proving to be perfectly suitable for run-time provisioning.

A comparison with utilization based state-of-the-art techniques and a rescaling algorithm shows that our solution outperform alternative methods

proving better results in terms of equilibrium efficiency both as regards the PoA both the IWC of each SaaS. In addition, our algorithm, achieve an efficient GNE under heavy workload conditions while thresholds based heuristic finds infeasible results due to the inability to manage SaaSs competition.

Finally analyses showed clear SaaS benefits while exploiting multiple IaaS deployment of applications and redistribution of traffic. SaaSs can have an average savings up to 50% compared to single IaaS architectures. The equilibria achieved are close to Cloud optimum, with inefficiency less than 2% for the social problem and of 6% for individual player in the worst cases.

Future work will extend the proposed solutions to consider multiple time-scales for performing resource allocation from few minutes to one hour. Additionally short-term solutions will be based on receding horizon techniques.

Bibliography

- [1] *2011 International Conference on Distributed Computing Systems, ICDCS 2011, Minneapolis, Minnesota, USA, June 20-24, 2011*. IEEE Computer Society, 2011.
- [2] *INFOCOM 2011. 30th IEEE International Conference on Computer Communications, Joint Conference of the IEEE Computer and Communications Societies, 10-15 April 2011, Shanghai, China*. IEEE, 2011.
- [3] *2012 IEEE 32nd International Conference on Distributed Computing Systems, Macau, China, June 18-21, 2012*. IEEE, 2012.
- [4] *26th IEEE International Parallel and Distributed Processing Symposium Workshops & PhD Forum, IPDPS 2012, Shanghai, China, May 21-25, 2012*. IEEE Computer Society, 2012.
- [5] Vineet Abhishek, Ian A. Kash, and Peter Key. Fixed and market pricing for cloud services. *CoRR*, abs/1201.5621, 2012.
- [6] A. Aleti, B. Buhnova, L. Grunske, A. Koziolk, and I. Meedeniya. Software architecture optimization methods: A systematic literature review.
- [7] Jussara Almeida, Virgílio Almeida, Danilo Ardagna, Ítalo Cunha, Chiara Francalanci, and Marco Trubian. Joint admission control and resource allocation in virtualized servers. *J. Parallel Distrib. Comput.*, 70(4):344–362, April 2010.
- [8] Eitan Altman, Thomas Boulogne, Rachid El Azouzi, Tania Jiménez, and Laura Wynter. A survey on networking games in telecommunications. *Computers & OR*, 33:286–311, 2006.
- [9] Amazon Inc. Amazon Elastic Cloud Computing. <http://aws.amazon.com/ec2/>.

BIBLIOGRAPHY

- [10] Amazon Inc. Amazon Web Services. <http://aws.amazon.com/>.
- [11] AMPL. Ampl modeling language for mathematical programming. <http://www.ampl.com/>.
- [12] Apache. Apache JMeter. <http://jmeter.apache.org/>.
- [13] D. Ardagna, E. di Nitto, P. Mohagheghi, S. Mosser, C. Ballagny, F. D'Andria, G. Casale, P. Matthews, C.-S. Nechifor, D. Petcu, A. Gericke, and C. Sheridan. ModacLOUDS: A model-driven approach for the design and execution of applications on multiple clouds. In *Modeling in Software Engineering (MISE), 2012 ICSE Workshop on*, pages 50–56, 2012.
- [14] Danilo Ardagna, Sara Casolari, Michele Colajanni, and Barbara Panicucci. Dual time-scale distributed capacity allocation and load redirect algorithms for cloud systems. *J. Parallel Distrib. Comput.*, 72(6):796–808, 2012.
- [15] Danilo Ardagna, Sara Casolari, and Barbara Panicucci. Flexible distributed capacity allocation and load redirect algorithms for cloud systems. In Liu and Parashar [58], pages 163–170.
- [16] Danilo Ardagna, Barbara Panicucci, and Mauro Passacantando. A game theoretic formulation of the service provisioning problem in cloud systems. In Srinivasan et al. [85], pages 177–186.
- [17] Danilo Ardagna, Barbara Panicucci, and Mauro Passacantando. Generalized nash equilibria for the service provisioning problem in cloud systems. *Services Computing, IEEE Transactions on*, PP(99):1, 2012.
- [18] Danilo Ardagna, Barbara Panicucci, Marco Trubian, and Li Zhang. Energy-aware autonomic resource allocation in multitier virtualized environments. *IEEE T. Services Computing*, 5(1):2–19, 2012.
- [19] Michael Armbrust, Michael Armbrust, Armando Fox, Rean Griffith, Anthony D. Joseph, Randy H. Katz, Andrew Konwinski, Gunho Lee, David A. Patterson, Ariel Rabkin, and Matei Zaharia. Above the clouds: A Berkeley view of cloud computing. 2009.
- [20] Kenneth Arrow and Gerard Debreu. Existence of an equilibrium for a competitive economy. *Econometrica*, 22(3):265–290, 1954.

-
- [21] Jeff Barr. *Host Your Web Site In The Cloud: Amazon Web Services Made Easy Amazon EC2 Made Easy*. Sitepoint, 1st edition, 2010.
- [22] Michael R Baye, Guoqiang Tian, and Jianxin Zhou. Characterizations of the existence of equilibria in games with discontinuous and non-quasiconcave payoffs. *Review of Economic Studies*, 60(4):935–48, October 1993.
- [23] R. Birke, L.Y. Chen, and E. Smirni. Data centers in the cloud: A large scale performance study. In *Cloud Computing (CLOUD), 2012 IEEE 5th International Conference on*, pages 336–343, June.
- [24] Mathias Björkqvist, Lydia Y. Chen, and Walter Binder. Opportunistic service provisioning in the cloud. In Chang [26], pages 237–244.
- [25] Marco Caldirola. Tecniche di resource allocation per sistemi virtualizzati di larga scala. Master’s thesis, Politecnico di Milano, 2010.
- [26] Rong Chang, editor. *2012 IEEE Fifth International Conference on Cloud Computing, Honolulu, HI, USA, June 24-29, 2012*. IEEE, 2012.
- [27] L. Cherkasova and P. Phaal. Session-based admission control: a mechanism for peak load management of commercial web sites. *Computers, IEEE Transactions on*, 51(6):669–685, Jun.
- [28] Gerard Debreu. A social equilibrium existence theorem. *Nat. Acad. Science*, 38:886–893, 1952.
- [29] Brian Dougherty, Jules White, and Douglas C. Schmidt. Model-driven auto-scaling of green cloud computing infrastructure. *Future Generation Comp. Syst.*, 28(2):371–378, 2012.
- [30] Parijat Dube, Zhen Liu, Laura Wynter, and Cathy H. Xia. Competitive equilibrium in e-commerce: Pricing and outsourcing. *Computers & OR*, 34(12):3541–3559, 2007.
- [31] Parijat Dube, Corinne Touati, and Laura Wynter. Capacity planning, quality of service and price wars. *SIGMETRICS Performance Evaluation Review*, 35(3):31–33, 2007.
- [32] Sourav Dutta, Sankalp Gera, Akshat Verma, and Balaji Viswanathan. Smartscale: Automatic application scaling in enterprise clouds. In Chang [26], pages 221–228.

BIBLIOGRAPHY

- [33] Yu.M. Ermoliev and S.D. Flaam. Repeated play of potential games. *Cybernetics and Systems Analysis*, 38:355–367, 2002.
- [34] Francisco Facchinei, Andreas Fischer, and Veronica Piccialli. Generalized nash equilibrium problems and newton methods. *Math. Program.*, 117(1-2):163–194, 2009.
- [35] Francisco Facchinei and Christian Kanzow. Generalized nash equilibrium problems. *Annals OR*, 175(1):177–211, 2010.
- [36] Francisco Facchinei, Veronica Piccialli, and Marco Sciandrone. Decomposition algorithms for generalized potential games. *Computational Optimization and Applications*, 50:237–262, 2011.
- [37] Yuan Feng, Baochun Li, and Bo Li. Price competition in an oligopoly cloud market. 2011.
- [38] Flexyscale. <http://www.flexyscale.com/>.
- [39] GoGrid. <http://www.gogrid.com/>.
- [40] Google Inc. <http://www.google.com/about/company/>.
- [41] Google Inc. Google App Engine. <https://developers.google.com/appengine/>.
- [42] Google Inc. Google Apps for Business. <http://www.google.com/enterprise/apps/business/>.
- [43] Google Inc. Google Compute Engine. <https://cloud.google.com/products/compute-engine>.
- [44] Hadi Goudarzi and Massoud Pedram. Multi-dimensional sla-based resource allocation for multi-tier cloud computing systems. In Liu and Parashar [58], pages 324–331.
- [45] Albert G. Greenberg and Kazem Sohraby, editors. *Proceedings of the IEEE INFOCOM 2012, Orlando, FL, USA, March 25-30, 2012*. IEEE, 2012.
- [46] Makhlof Hadji and Djamel Zeghlache. Minimum cost maximum flow algorithm for dynamic resource allocation in clouds. In Chang [26], pages 876–882.

-
- [47] Mohammad Mehedi Hassan, M.Shamim Hossain, A.M.Jehad Sarkar, and Eui-Nam Huh. Cooperative game-based distributed resource allocation in horizontal dynamic cloud federation platform. *Information Systems Frontiers*, pages 1–20, 2012.
- [48] Mohammad Mehedi Hassan, Biao Song, and Eui nam Huh. Distributed resource allocation games in horizontal dynamic cloud federation platform. In Thulasiraman et al. [87], pages 822–827.
- [49] Ting He, Shiyao Chen, Hyoil Kim, Lang Tong, and Kang-Won Lee. Scheduling parallel tasks onto opportunistically available cloud resources. In Chang [26], pages 180–187.
- [50] IBM. IBM ILOG CPLEX Optimizer. <http://www-01.ibm.com/software/integration/optimization/cplex-optimizer/>.
- [51] Tatsuuro Ichiishi. *Game Theory for Economic Analysis*. New York: Academic Press, 1983.
- [52] Ganesh Neelakanta Iyer and Bharadwaj Veeravalli. On the resource allocation and pricing strategies in compute clouds using bargaining approaches. In Veeravalli and Foster [91], pages 147–152.
- [53] JMeter Plugins. <https://code.google.com/p/jmeter-plugins/>.
- [54] Kernel Based Virtual Machine. <http://www.linux-kvm.org/>.
- [55] Kleopatra Konstanteli, Tommaso Cucinotta, Konstantinos Psychas, and Theodora A. Varvarigou. Admission control for elastic cloud services. In Chang [26], pages 41–48.
- [56] Dara Kusic, Jeffrey O. Kephart, James E. Hanson, Nagarajan Kandasamy, and Guofei Jiang. Power and performance management of virtualized computing environments via lookahead control. *Cluster Computing*, 12(1):1–15, 2009.
- [57] Yi-Kuei Lin and Ping-Chen Chang. Reliability evaluation of a computer network in cloud computing environment subject to maintenance budget. *Applied Mathematics and Computation*, 219(8):3893–3902, 2012.
- [58] Ling Liu and Manish Parashar, editors. *IEEE International Conference on Cloud Computing, CLOUD 2011, Washington, DC, USA, 4-9 July, 2011*. IEEE, 2011.

BIBLIOGRAPHY

- [59] Tieming Liu, Chinnatat Methapatara, and Laura Wynter. Revenue management model for on-demand it services. *European Journal of Operational Research*, 207(1):401–408, 2010.
- [60] Markov4JMeter. <http://se.informatik.uni-kiel.de/markov4jmeter/>.
- [61] Michele Mazzucco and Dmytro Dyachuk. Optimizing cloud providers revenues via energy efficient server allocation. *Sustainable Computing: Informatics and Systems*, 2(1):1 – 12, 2012.
- [62] Peter Mell and Tim Grance. The NIST Definition of Cloud Computing. Technical report, July 2009.
- [63] Ishai Menache, Asuman Ozdaglar, and Nahum Shimkin. Socially optimal pricing of cloud computing resources. In *Proceedings of the 5th International ICST Conference on Performance Evaluation Methodologies and Tools, VALUETOOLS '11*, pages 322–331, ICST, Brussels, Belgium, Belgium, 2011. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).
- [64] Microsoft Corporation. <http://www.microsoft.com/>.
- [65] Microsoft Corporation. Microsoft Office 365. <http://office365.microsoft.com/>.
- [66] Microsoft Corporation. Microsoft Windows Azure. <http://www.windowsazure.com/>.
- [67] Microsoft Corporation. Microsoft Windows Azure Virtual Machines. <http://www.windowsazure.com/en-us/home/features/virtual-machines/>.
- [68] MODAClouds. <http://www.modaclouds.eu/>.
- [69] Dov Monderer and Lloyd S. Shapley. Potential games. *Games and Economic Behavior*, 14(1):124–143, May 1996.
- [70] Jacqueline Morgan and Vincenzo Scalzo. Pseudocontinuous functions and existence of nash equilibria. *Journal of Mathematical Economics*, 43(2):174 – 183, 2007.
- [71] John Nash. Non-cooperative games. *The Annals of Mathematics*, 54(2):286–295, 1951.

-
- [72] Elisabetta Di Nitto and Ramin Yahyapour, editors. *Towards a Service-Based Internet - Third European Conference, ServiceWave 2010, Ghent, Belgium, December 13-15, 2010. Proceedings*, volume 6481 of *Lecture Notes in Computer Science*. Springer, 2010.
- [73] C.G.A.M. van den Nouweland, Peter Borm, W. van Golstein Brouwers, R. Groot Bruinderink, and S.H. Tijs. A game theoretic approach to problems in telecommunication. Open access publications from tilburg university, Tilburg University, 1996.
- [74] Onlive. <http://www.onlive.com/>.
- [75] Rackspace Inc. <http://www.rackspace.com/>.
- [76] N.S.V. Rao, S.W. Poole, Fei He, Jun Zhuang, C.Y.T. Ma, and D.K.Y. Yau. Cloud computing infrastructure robustness: A game theory approach. In *Computing, Networking and Communications (ICNC), 2012 International Conference on*, pages 34–38, 30 2012-feb. 2 2012.
- [77] Philip J. Reny. On the existence of pure and mixed strategy nash equilibria in discontinuous games. *Econometrica*, 67(5):1029–1056, 1999.
- [78] Salesforce Inc. <http://www.force.com/>.
- [79] SAP. <http://www.sap.com/>.
- [80] U. Sharma, P. Shenoy, S. Sahu, and A. Shaikh. A cost-aware elasticity provisioning system for the cloud. In *Distributed Computing Systems (ICDCS), 2011 31st International Conference on*, pages 559–570, june 2011.
- [81] Upendra Sharma, Prashant J. Shenoy, Sambit Sahu, and Anees Shaikh. Kingfisher: Cost-aware elasticity in the cloud. In *INFOCOM [2]*, pages 206–210.
- [82] Socialbakers. Facebook statistics. <http://www.socialbakers.com/facebook-statistics/>.
- [83] Yang Song, Murtaza Zafer, and Kang-Won Lee. Optimal bidding in spot instance market. In Greenberg and Sohraby [45], pages 190–198.
- [84] Shekhar Srikantaiah, Aman Kansal, and Feng Zhao. Energy aware consolidation for cloud computing. In *Proceedings of the 2008 conference on Power aware computing and systems, HotPower’08*, pages 10–10, Berkeley, CA, USA, 2008. USENIX Association.

BIBLIOGRAPHY

- [85] Sadagopan Srinivasan, Krithi Ramamritham, Arun Kumar, M. P. Ravindra, Elisa Bertino, and Ravi Kumar, editors. *Proceedings of the 20th International Conference on World Wide Web, WWW 2011, Hyderabad, India, March 28 - April 1, 2011*. ACM, 2011.
- [86] Standard Performance Evaluation Corporation. SPECweb2005. <http://www.spec.org/web2005/>.
- [87] Parimala Thulasiraman, Laurence Tianruo Yang, Qiwen Pan, Xingang Liu, Yaw-Chung Chen, Yo-Ping Huang, Lin-Huang Chang, Che-Lun Hung, Che-Rung Lee, Justin Y. Shi, and Ying Zhang, editors. *13th IEEE International Conference on High Performance Computing & Communication, HPCC 2011, Banff, Alberta, Canada, September 2-4, 2011*. IEEE, 2011.
- [88] Fengguang Tian and Keke Chen. Towards optimal resource provisioning for running mapreduce programs in public clouds. In Liu and Parashar [58], pages 155–162.
- [89] Guoqiang Tian and Jianxin Zhou. Transfer continuities, generalizations of the weierstrass and maximum theorems: A full characterization. *Journal of Mathematical Economics*, 24(3):281 – 303, 1995.
- [90] Luis M. Vaquero, Luis Rodero-Merino, Juan Caceres, and Maik Lindner. A break in the clouds: towards a cloud definition. *SIGCOMM Comput. Commun. Rev.*, 39(1):50–55, December 2008.
- [91] Bharadwaj Veeravalli and Ian T. Foster, editors. *Proceedings of the 17th IEEE International Conference on Networks, ICON 2011, Singapore, December 14-16, 2011*. IEEE, 2011.
- [92] VMware Inc. <http://www.vmware.com/>.
- [93] Mark Voorneveld. Equilibria and approximate equilibria in infinite potential games. *Economics Letters*, 56(2):163 – 169, 1997.
- [94] Jian Wan, Dechuan Deng, and Congfeng Jiang. Non-cooperative gaming and bidding model based resource allocation in virtual machine environment. In *IPDPS Workshops* [4], pages 2183–2188.
- [95] Hongyi Wang, Qingfeng Jing, Rishan Chen, Bingsheng He, Zhengping Qian, and Lidong Zhou. Distributed systems meet economics: pricing in the cloud. In *Proceedings of the 2nd USENIX conference on Hot topics in cloud computing*, HotCloud’10, pages 6–6, Berkeley, CA, USA, 2010. USENIX Association.

-
- [96] J Wardrop. Some theoretical aspects of road traffic research. *Proceedings of the Institution of Civil Engineers, Part II*, 1(36):352–362, 1952.
- [97] Guiyi Wei, Athanasios V. Vasilakos, Yao Zheng, and Naixue Xiong. A game-theoretic method of fair resource allocation for cloud computing services. *The Journal of Supercomputing*, 54(2):252–269, 2010.
- [98] Andreas Wolke and Gerhard Meixner. Twospot: A cloud platform for scaling out web applications dynamically. In Nitto and Yahyapour [72], pages 13–24.
- [99] Linlin Wu, Saurabh Kumar Garg, and Rajkumar Buyya. Sla-based admission control for a software-as-a-service provider in cloud computing environments. *J. Comput. Syst. Sci.*, 78(5):1280–1299, 2012.
- [100] Xen. Xen Hypervisor. <http://www.xen.org/>.
- [101] Z. Xiao, Q. Chen, and H. Luo. Automatic scaling of internet applications for cloud computing services. *Computers, IEEE Transactions on*, PP(99):1, 2012.
- [102] Z. Xiao, W. Song, and Q. Chen. Dynamic resource allocation using virtual machines for cloud computing environment. *Parallel and Distributed Systems, IEEE Transactions on*, PP(99):1, 2012.
- [103] PengCheng Xiong, Zhikui Wang, Simon Malkowski, Qingyang Wang, Deepal Jayasinghe, and Calton Pu. Economical and robust provisioning of n-tier cloud workloads: A multi-level control approach. In *ICDCS* [1], pages 571–580.
- [104] Murtaza Zafer, Yang Song, and Kang-Won Lee. Optimal bids for spot vms in a cloud for deadline constrained jobs. In Chang [26], pages 75–82.
- [105] Sharrukh Zaman and Daniel Grosu. An online mechanism for dynamic vm provisioning and allocation in clouds. In Chang [26], pages 253–260.
- [106] Qi Zhang, Lu Cheng, and Raouf Boutaba. Cloud computing: state-of-the-art and research challenges. *J. Internet Services and Applications*, 1(1):7–18, 2010.
- [107] Qi Zhang, Quanyan Zhu, Mohamed Faten Zhani, and Raouf Boutaba. Dynamic service placement in geographically distributed clouds. In *ICDCS* [3], pages 526–535.

BIBLIOGRAPHY

- [108] Xiaoyun Zhu, Donald Young, Brian J. Watson, Zhikui Wang, Jerry Rolia, Sharad Singhal, Bret McKee, Chris Hyser, Daniel Gmach, Rob Gardner, Tom Christian, and Ludmila Cherkasova. 1000 islands: an integrated approach to resource management for virtualized data centers. *Cluster Computing*, 12(1):45–57, 2009.