References

[1] Michael Armbrust, Michael Armbrust, A Fox, A Fox, R Griffith, R Griffith, AD Joseph, AD Joseph, RH, and RH. Above the clouds: A Berkeley view of cloud computing. *University of California, Berkeley, Tech. Rep. UCB*, pages 07–013, 2009.

[2] Amazon ec2 instance types. `http://aws.amazon.com/ec2/instance-types`. Accessed: 2015-04-01.

[3] Amazon web services - auto scaling. `http://aws.amazon.com/autoscaling/`. Accessed: 2014-10-20.

[4] Tania Lorido-Botran, Jose Miguel Alonso, and Jose A. Lozano. A Review of Auto-scaling Techniques for Elastic Applications in Cloud Environments. *Journal of Grid Computing*, (1):1–34, 2014.

[5] Jason Mars, Lingjia Tang, Robert Hundt, Kevin Skadron, and Mary Lou Soffa. Bubble-Up. In *In Proc. of the Annual IEEE/ACM International Symposium on Microarchitecture*, page 248, 2011.

[6] Rui Han, Li Guo, Moustafa Ghanem, and Yike Guo. Lightweight resource scaling for cloud applications. In *CCGRID*, pages 644–651. IEEE, 2012.

[7] Masum Z. Hasan, Edgar Magana, Alexander Clemm, Lew Tucker, and Sree Lakshmi D. Gudreddi. Integrated and autonomic cloud resource scaling. In *NOMS*, pages 1327–1334. IEEE, 2012.

[8] Michael Maurer, Ivona Brandic, and Rizos Sakellariou. Enacting slas in clouds using rules. In Emmanuel Jeannot, Raymond Namyst, and Jean Roman, editors, *Euro-Par (1)*, volume 6852 of *Lecture Notes in Computer Science*, pages 455–466. Springer, 2011.

[9] Auto scaling in the amazon cloud. `http://techblog.netflix.com/2012/01/auto-scaling-in-amazon-cloud.html`. Accessed: 2015-03-20.

[10] Peter Bodík, Rean Griffith, Charles A. Sutton, Armando Fox, Michael I. Jordan, and David A. Patterson. Statistical machine learning makes automatic control practical for internet datacenters. In *Workshop on Hot Topics in Cloud Computing, HotCloud'09, San Diego, CA, USA, June 15, 2009*, 2009.

[11] Enda Barrett, Enda Howley, and Jim Duggan. Applying reinforcement learning towards automating resource allocation and application scalability in the cloud. *Concurrency and Computation: Practice and Experience*, 25(12):1656–1674, 2013.

[12] Xavier Dutreilh, Sergey Kirgizov, Olga Melekhova, Jacques Malenfant, Nicolas Rivierre, and Isis Truck. Using Reinforcement Learning for Autonomic Resource Allocation in Clouds: towards a fully automated workflow. In *Seventh International Conference on Autonomic and Autonomous Systems, ICAS 2011*, pages 67–74. IEEE, May 2011. MoVe INT LIP6.

[13] Jia Rao, Xiangping Bu, Cheng-Zhong Xu, Le Yi Wang, and Gang George Yin. Vconf: a reinforcement learning approach to virtual machines auto-configuration. In Simon A. Dobson, John Strassner, Manish Parashar, and Onn Shehory, editors, *ICAC*, pages 137–146. ACM, 2009.

[14] Xavier Dutreilh and Sergey Kirgizov. Using reinforcement learning for autonomic resource allocation in clouds: towards a fully automated workflow. *International Conference on Autonomic and Autonomous Systems*, (c):67–74, 2011.

[15] Jia Rao, Xiangping Bu, Cheng zhong Xu, Leyi Wang, and George Yin. Vconf: a reinforcement learning approach to virtual machines auto-configuration. In *In ICAC*, 2009.

[16] Harold C. Lim, Shivnath Babu, Jeffrey S. Chase, and Sujay S. Parekh. Automated control in cloud computing: Challenges and opportunities. In *In First Workshop on Automated Control for Datacenters and Clouds*, 2009.

[17] Harold Lim, Shivnath Babu, and Jeffrey S. Chase. Automated control for elastic storage. In Manish Parashar, Renato J. Figueiredo, and Emre Kiciman, editors, *ICAC*, pages 1–10. ACM, 2010.

[18] Bhuvan Urgaonkar, Prashant J. Shenoy, Abhishek Chandra, Pawan Goyal, and Timothy Wood. Agile dynamic provisioning of multi-tier internet applications. *TAAS*, 3(1), 2008.

[19] Daniel A. Villela, Prashant Pradhan, and Dan Rubenstein. Provisioning servers in the application tier for e-commerce systems. *ACM Trans. Internet Techn.*, 7(1), 2007.

[20] Hien Nguyen Van, Frédéric Dang Tran, and Jean-Marc Menaud. Sla-aware virtual resource management for cloud infrastructures. In *International Conference on Computer and Information Technology, Xiamen, China*, pages 357–362, 2009.

[21] Makhlouf Hadji and Djamal Zeghlache. Minimum cost maximum flow algorithm for dynamic resource allocation in clouds. In *International Conference on Cloud Computing, Honolulu, HI, USA*, pages 876–882, 2012.

[22] Norman Bobroff, Andrzej Kochut, and Kirk Beaty. Dynamic placement of virtual machines for managing sla violations. In *Integrated Network Management*, pages 119–128. IEEE, 2007.

[23] Nicolo Maria Calcavecchia, Ofer Biran, Erez Hadad, and Yosef Moatti. VM Placement Strategies for Cloud Scenarios. *International Conference on Cloud Computing*, pages 852–859, June 2012.

[24] Emiliano Casalicchio, Daniel a. Menascé, and Arwa Aldhalaan. Autonomic resource provisioning in cloud systems with availability goals. *Proceedings of*

the 2013 ACM Cloud and Autonomic Computing Conference on - CAC '13, page 1, 2013.

[25] Xiaoqiao Meng, Canturk Isci, Jeffrey Kephart, Li Zhang, Eric Bouillet, and Dimitrios Pendarakis. Efficient resource provisioning in compute clouds via VM multiplexing. *Proceeding of the International conference on Autonomic computing*, page 11, 2010.

[26] Ajay Gulati, Ganesha Shanmuganathan, Anne M. Holler, and Irfan Ahmad. Cloud scale resource management: Challenges and techniques. In *USENIX Workshop on Hot Topics in Cloud Computing Portland, OR, USA*, 2011.

[27] David Erickson, Brandon Heller, Nick McKeown, and Mendel Rosenblum. Using network knowledge to improve workload performance in virtualized data centers. In *International Conference on Cloud Engineering, Boston, MA, USA*, pages 185–194, 2014.

[28] Chunqiang Tang, Malgorzata Steinder, Michael Spreitzer, and Giovanni Pacifici. A Scalable Application Placement Controller for Enterprise Data Centers. *Proceedings of the 16th international conference on World Wide Web - WWW '07*, page 331, 2007.

[29] Ron Chiang, Jinho Hwang, Howie Huang, and Timothy Wood. Matrix: Achieving predictable virtual machine performance in the clouds. In *ICAC*. IEEE Computer Society, 2014.

[30] Brian J. Watson, Manish Marwah, Daniel Gmach, Yuan Chen, Martin Arlitt, and Zhikui Wang. Probabilistic performance modeling of virtualized resource allocation. *Proceeding of the International conference on Autonomic computing*, page 99, 2010.

[31] Archana Ganapathi, Harumi Kuno, Umeshwar Dayal, Janet L. Wiener, Armando Fox, Michael Jordan, and David Patterson. Predicting multiple metrics for queries: Better decisions enabled by machine learning. In

Proceedings - International Conference on Data Engineering, pages 592–603, 2009.

[32] Sajib Kundu, Raju Rangaswami, Ajay Gulati, Ming Zhao, and Kaushik Dutta. Modeling virtualized applications using machine learning techniques. *Proc. of the ACM SIGPLAN/SIGOPS conference on Virtual Execution Environments*, page 3, 2012.

[33] Rubis: Rice university bidding system. `http://rubis.ow2.org/`. Accessed: 2014-08-05.

[34] Filebench: a framework for simulating applications on file systems. `http://www.solarisinternals.com/wiki/index.php/FileBench`. Accessed: 2015-05-20.

[35] Nilabja Roy, Abhishek Dubey, and Aniruddha Gokhale. Efficient autoscaling in the cloud using predictive models for workload forecasting. In *Proc. of the International Conference on Cloud Computing*, pages 500–507, 2011.

[36] Performing vm migration under xen. `http://wiki.xenproject.org/wiki/Migration`. Accessed: 2015-03-16.

[37] Vmware vsphere. `https://www.vmware.com/products/vsphere/features/vmotion`. Accessed: 2015-03-16.

[38] Akshat Verma, Gautam Kumar, and Ricardo Koller. The cost of reconfiguration in a cloud. In *Proceedings of the 11th International Middleware Conference Industrial track*, pages 11–16. ACM, 2010.

[39] William Voorsluys, James Broberg, Srikumar Venugopal, and Rajkumar Buyya. Cost of virtual machine live migration in clouds: A performance evaluation. *Lecture Notes in Computer Science*, 5931 LNCS:254–265, 2009.

[40] David Breitgand, Gilad Kutiel, and Danny Raz. Cost-aware live migration of services in the cloud. *Proc. of the ACM International Systems & Storage Conference*, page 1, 2010.

[41] Guofu Feng, Saurabh Garg, Rajkumar Buyya, and Wenzhong Li. Revenue Maximization Using Adaptive Resource Provisioning in Cloud Computing Environments. *2012 ACM/IEEE 13th International Conference on Grid Computing*, pages 192–200, September 2012.

[42] Li Zhang and Danilo Ardagna. Sla based profit optimization in web systems. In *Proceedings of the 13th international World Wide Web conference on Alternate track papers & posters*, pages 462–463. ACM, 2004.

[43] Rodrigo N. Calheiros, Rajiv Ranjan, CÃľsar A. F. De Rose, and Rajkumar Buyya. Cloudsim: A novel framework for modeling and simulation of cloud computing infrastructures and services. *CoRR*, abs/0903.2525, 2009.

[44] Marc Bux and Ulf Leser. Dynamiccloudsim: Simulating heterogeneity in computational clouds. In *Proceedings of the 2Nd ACM SIGMOD Workshop on Scalable Workflow Execution Engines and Technologies*, SWEET '13, pages 1:1–1:12, New York, NY, USA, 2013. ACM.

[45] Nikolay Grozev and Rajkumar Buyya. Performance Modelling and Simulation of Three-Tier Applications in Cloud and Multi-Cloud Environments. *The Computer Journal*, pages bxt107–, 2013.

[46] Kullback-liebler divergence. `https://en.wikipedia.org/wiki/Kullback-Leibler_divergence`. Accessed: 2015-06-01.

[47] Haikun Liu, Cheng-Zhong Xu, Hai Jin, Jiayu Gong, and Xiaofei Liao. Performance and energy modeling for live migration of virtual machines. In *HPDC*, pages 171–182. ACM, 2011.

[48] Felix Salfner, Peter Tr, and Andreas Polze. Downtime Analysis of Virtual Machine Live Migration. *International Conference on Dependability*, (c):100–105, 2011.

[49] Ming Mao and Marty Humphrey. A performance study on the vm startup time in the cloud. In Rong Chang, editor, *IEEE CLOUD*, pages 423–430, 2012.

[50] libvirt virtualization api. http://libvirt.org/. Accessed: 2014-07-25.

[51] Rubis workload: Simple installation guide. http://sanifool.com/2012/09/03/rubis-workload-simple-installation-guide. Accessed: 2014-11-10.

[52] Aaron Beitch, Brandon Liu, Timothy Yung, Rean Griffith, Armando Fox, and David Patterson. Rain: A workload generation toolkit for cloud computing applications. *Technical Report No. UCB/EECS-2010-14*, 2010.

[53] Marco Guazzone. The rubis workload implementation for rain. https://github.com/sguazt/rain-workload-toolkit, 2013. Accessed: 2015-01-22.

[54] David Mosberger and Tai Jin. Httperf—A Tool for Measuring Web Server Performance. *ACM SIGMETRICS Performance Evaluation Review*, 26(3):31–37, December 1998.

[55] Raoufehsadat Hashemian, Diwakar Krishnamurthy, and Martin Arlitt. Web workload generation challenges–an empirical investigation. *Software: Practice and Experience*, 42(5):629–647, 2012.

[56] httperf patch. http://people.ucalgary.ca/~dkrishna/SPE. Accessed: 2014-11-21.

[57] httperf patch. https://github.com/klueska/httperf. Accessed: 2015-02-30.

[58] Linux: Increasing the number of open file descriptors. https://cs.uwaterloo.ca/~brecht/servers/openfiles.html. Accessed: 2015-02-22.

[59] sysstat. http://sebastien.godard.pagesperso-orange.fr. Accessed: 2014-07-11.

[60] Worldcup 98. http://ita.ee.lbl.gov/html/contrib/WorldCup.html. Accessed: 2014-12-10.

[61] fio: Flexible i/o tester. http://freshmeat.net/projects/fio. Accessed: 2015-05-15.

[62] How linkedin gets twenty times more money per user than facebook. http://www.businessinsider.com/linked-revenue-facebook-2012-7. Accessed: 2015-07-12.