Economics-inspired Resource and Energy Management for Cloud Environments (Keynote on ESaaSA Workshop CLOSER 2015)

Luís Veiga
INESC-ID Lisboa
Instituto Superior Técnico
Universidade de Lisboa

May 2015

Abstract

Cloud infrastructures host different kinds of applications belonging to clients with different levels of service agreements. Their execution is supported by high language virtual machines and system-level virtual machines (VMs). Aiming to maximize revenue, by minimizing the operational expenditure, cloud providers often consolidate several VMs in a single server. This is particularly useful also for the emergent distributed clouds where physical resources, at each node, are not so abundant. However, this technique can lead to overcommitment of resources, and to undesirable performance degradation, if carried out in a non-informed way.

Our recent work proposes new allocation mechanisms for the VMs used by Platform-as-a-Service (PaaS) and Infrastructure-as-a-Service (IaaS) clouds, and how these mechanisms are to be controlled. They are driven by classic economic notions such as yield, which expresses the return the provider has in applying a given resource allocation to the tenants' workload, and utility functions, a relation of clients' perceived usefulness to a given allocation.

For PaaS providers, a Java VM was extended with an integrated resource management API, heap resizing policies for yield maximization and concurrent checkpoint for migration of the execution state. Overall, these new mechanisms impose small penalties, measured in the execution of typical benchmarks, while allowing the use of application-tailored policies.

At the IaaS level, we present a novel cost model and new scheduling algorithms for system-level VMs, along with their implementation in a state-of-the-art cloud simulation framework. Simulations with synthetic and real-world traces, show that the utility-based scheduling allows more VMs to be allocated, thus allowing extra revenue per resource allocated, and shorter waiting times for clients, when comparing with a utility-oblivious redistribution of resources.

Scheduling algorithms are also extended to take into account energy efficiency. For this assessment the typical data center topologies are compared against community networks and community cloud scenarios, which are backed by peer-to-peer architectures.

http://www.gsd.inesc-id.pt/~lveiga/papers/keynote-CLOSER-2015-ESaaSA.pdf

References:

[J1] José Simão and Luís Veiga, Partial Utility-driven Scheduling for Flexible SLA and Pricing Arbitration in Cloud, IEEE Transactions on Cloud Computing, Nov. 2014, IEEE, online first

[J2] José Simão and Luís Veiga, Adaptability Driven by Quality Of Execution in High Level Virtual Machines for Shared Environments, International Journal of Computer Systems Science and Engineering, Nov. 2013, CRL publishing.

[J3] José Simão et al. , A Checkpointing-enabled and Resource-Aware Java VM for Efficient and Robust e-Science Applications in Grid Environments, Concurrency and Computation: Practice and Experience, Sep. 2012, Wiley

[C1] José Simão and Luís Veiga, Flexible SLAs in the Cloud with Partial Utility-driven Scheduling, IEEE 5th International Conference on Cloud Computing Technology and Science (CloudCom 2013), Dec. 2013, IEEE. (Best-Paper Award Runner-up)

[C2] José Simão et.al., A2-VM: A Cooperative Java VM with Support for Resource-Awareness and Cluster-Wide Thread Scheduling, 19th

International Conference on COOPERATIVE INFORMATION SYSTEMS (CoopIS 2011), Sep. 2011, Springer
[C3] José Simão and Luís Veiga, A Progress and Profile-driven Cloud-VM for Improved Resource-Efficiency and Fairness in e-Science Environments, 28th ACM Symposium On Applied Computing (SAC 2013), Mar. 2013, ACM
[C4] João Pedro Silva and José Simão and Luís Veiga, Ditto – Deterministic Execution Replayability-as-a-Service for Java VM on Multiprocessors, ACM/IFIP/Usenix International Middleware Conference (Middleware 2013), Dec. 2013, Springer.
[C5] Leila Sharifi et al. Energy Efficiency Dilemma: P2P-Cloud vs. Datacenter, IEEE 6th International Conference on Cloud Computing Technology and Science (CloudCom 2014), Dec. 2014 IEEE. (Best Paper Candidate)
[W1] José Simão and Luís Veiga, A Classification of Middleware to Support Virtual Machines Adaptability in IaaS, 11th International Workshop on Adaptive and Reflective Middleware (ARM 2012), In conjuntion with Middleware 2012.
[W2] José Simão and Luís Veiga, VM Economics for Java Cloud Computing - An Adaptive and Resource-Aware Java Runtime with Quality-of-Execution, CCGrid 2012 - Doctoral Symposium: Cloud Scheduling, Clusters and Data Centers, May 2012, IEEE

Acknowledgements: PhD and MSc students: José Simão (recently graduated Ph.D.) and Leila Sharifi (graduating Ph.D. 2016)