Lifting cloud infrastructure service consumers to providers

Josef Spillner[1] and Andrey Brito[2]
[1]Technische Universität Dresden, Faculty of Computer Science,
01062 Dresden, Germany
Email: `josef.spillner@tu-dresden.de`
[2]Universidade Federal de Campina Grande, Departamento de Sistemas e Computação,
Laboratório de Sistemas Distribuídos,
CEP 58429-900, Campina Grande - PB, Brazil
Email: `andrey@dsc.ufcg.edu.br`

ABSTRACT

Today's commercial cloud infrastructure service providers offer limited resource elasticity, leaving a gap between the vision of full on-demand/utility computing ("computation like electricity from the power plug") and the economically or technically motivated limitations of virtualisers, cloud stacks and services. For instance, Amazon as popular resource vendor mandates a lower elasticity boundary of 1 hour of computation (EC2 and ELB), 1 GB of data storage (EBS and S3), 1 GB of transfer (EC2 and ELB), and a maximum of 20 virtual machines unless individually negotiated otherwise. Furthermore, the scalability is coarse-grained both horizontally and vertically, because the amount of resources is a multiple of an instance type of which only a few exist. This makes it hard to implement dynamic job scheduling and data partitioning algorithms for processing and storage in the cloud without wasting resources, and consequently money, due to the constrained allocation policies. The problem is made worse by the lack of formal cloud provider descriptions. Hence, users migrating between providers or aggregating resources need to manually keep track of these limits, which are either informally specified in the documentation and pricing terms, or even not at all. With the emerging nested virtualisation capabilities in modern operating systems [1, 2], it is possible to let consumers run their own nested machines inside the virtual machines. Using service management platforms, they then offer these machines through markets to prospective sub-consumers in need of short-term peak capacities or long-term micro capacities, who can again, recursively, repurpose them.

BODY

Nested virtualisation is a low-overhead mechanism to partially repurpose leased compute resources. This leads to recursive cloud providers.

REFERENCES

[1] M. Ben-Yehuda, M. D. Day, Z. Dubitzky, M. Factor, N. Har'El, A. Gordon, A. Liguori, O. Wasserman, and B.-A. Yassour. The turtles project: Design and implementation of nested virtualization. In *Proceedings of the 9th USENIX Conference on Operating Systems Design and Implementation (OSDI)*, pages 423–436, October 2010. Vancouver, British Columbia, Canada.

[2] Z. Pan, Q. He, W. Jiang, Y. Chen, and Y. Dong. Nestcloud: Towards practical nested virtualization. In *International Conference on Cloud and Service Computing (CSC)*, pages 321–329, December 2011. Hong Kong, China.