

Economics of a Supercloud

Ian Kash[†] Qin Jia Zhiming Shen Weijia Song Robbert van Renesse Hakim Weatherspoon

Microsoft Research[†] Cornell University

1. Introduction

A Supercloud is a “CrossCloud”: It is an Infrastructure-as-a-Service (IaaS) that goes beyond federated or hybrid clouds and gives its users direct control over cloud deployments—even across different underlying cloud providers (Jia et al. 2015). It supports privileged cloud operations such as migration across autonomous cloud providers even if they use different virtual machine monitors, networking, and storage infrastructure. For instance, a user can start a virtual machine instance in the Google cloud (Google Compute Engine), migrate it live to the Amazon cloud (Amazon Elastic Compute Cloud), live migrate it again to the Microsoft cloud (Microsoft Azure), then finally live migrate it to a private cloud (e.g. a Eucalyptus or OpenStack based cloud). Moreover, a user can have a deployment that simultaneously exists over various combination of cloud providers, then change the placement at any time. To support such privileged operations and customizations across many different cloud providers, the Supercloud provides virtual machine, storage, and networking complete with a full set of management operations, allowing applications to optimize performance across different underlying cloud providers.

Given the new capability that a Supercloud offers, in this paper we examine both the economic advantages of a Supercloud and its implications for cloud providers. In particular, we examine five different channels by which the Supercloud provides benefits to its users

1. Pricing power
2. Geographic mobility
3. Increased utilization
4. Spot market participation
5. Smooth out stochastic demands

For each of these channels we discuss how they benefit the users of a Supercloud, what effect this use of the Supercloud has on the cloud provider (who has different economic interests from the cloud users), and related research questions.

2. Supercloud Economics

2.1 Pricing Power

Cloud providers invest in datacenters and hardware well before they begin to use the resulting capacity, and these capital investments have a lifetime of years. Having invested in this capital, they then compete with each other to profit from their investments. This makes the cloud a classic example of *Cournot competition*, where firms first determine how much of their product to produce and

then set prices to sell it. Thus, even without the Supercloud, IaaS cloud is a competitive marketplace, and prices are largely similar across providers for basic offerings (for example, Microsoft Azure has committed to matching a number of Amazon’s prices (pri a,b)).

Effect on Users Since prices are already competitive, using a Supercloud is unlikely to directly lead to better prices. However, providers package resources slightly differently and Supercloud users can selectively and fluidly choose different packaging to fit their workloads.

Effect on Providers One feature that the Cournot competition model does not capture is that it is currently difficult and expensive for a company (cloud user) to move its whole operation from one cloud to another, which gives cloud providers pricing power over their existing customers. The threat of a Supercloud undoes the pricing power and may well lower prices somewhat as an indirect effect for all customers, not necessarily just ones using a Supercloud.

Research Questions To what extent is the model of a Cournot competition adequate to capture the economics of the cloud? Richer models explored thus far can lead to non-uniform pricing due to differences in quality (Kilcioglu and Rao 2016; Anselmi et al. 2014), but the basic message seems to be the same.

2.2 Geographic Mobility

The Supercloud allows live migration of resources (virtual machine instances, storage, and networking) from one datacenter to another, and from one cloud provider to another, which allows resources to be moved to where the load is high. As a result of the geographic mobility enabled by the Supercloud, customers can “follow the sun” (Jia et al. 2015) and move (stateful) resources to where they are needed.

Effect on Users Geographic mobility allows customers to “auto scale” across different locations and different providers and, in aggregate, rent fewer resources: Resources can be concentrated to respond to load.

Effect on Providers On one hand geographic mobility means providers sell less to each customer, but on the other this capability can make the cloud more attractive to customers, so the effects are ambiguous.

Research Questions How should cloud providers trade off efficiency and revenue? Similar questions have been explored in advertising auctions (Bachrach et al. 2014).

2.3 Increased Utilization

Similarly, a Supercloud user can increase the density of virtual machines handled by a physical machine at times of low load. Therefore, we expect its effects to be broadly similar to geographic mobility. However, unlike geographic mobility, which requires geo-replication of data and state and can be financially expensive to migrate across cloud providers or across availability zones, increasing utilization within an availability zone avoids the financial burdens of these types of migrations. Instead, the Supercloud enables migrating to different types of virtual machine instance (e.g. from 2XL to 4XL instances) (Jia et al. 2016).

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, or to redistribute to lists, contact the Owner/Author(s). Request permissions from permissions@acm.org or Publications Dept., ACM, Inc., fax +1 (212) 869-0481.

CrossCloud'16 April 18, 2016, London, United Kingdom
Copyright © 2016 held by owner/author(s). Publication rights licensed to ACM.
ACM 978-1-4503-4294-0/16/04...\$15.00
DOI: <http://dx.doi.org/10.1145/http://dx.doi.org/10.1145/2904111.2904119>

Effect on Users Switching to lower cost instances clearly saves customers money.

Effect on Providers Provisioning for instances that might eventually use their resources but currently are not is inefficient and expensive for the cloud provider, so this ease of migration may be beneficial even if customers pay lower costs.

Research Questions Is this something that a Supercloud is naturally better at than a cloud provider, or is it a feature that competition will force providers to start doing themselves?

2.4 Spot Markets

The Amazon spot market and Google preemptable services offer reduced pricing for virtual machine instances if a user is willing to tolerate an instance termination at any time. The Supercloud can avoid termination by rapidly transitioning to, e.g. a standard instance.

Effect on Users In the short term, such spot market arbitrage would be beneficial for Supercloud customers. That is, they can make use of cheap spot instances when they are available and standard instances the rest of the time, while achieving the same reliability as if they were using standard instances the whole time.

Effect on Providers Amazon and Google are willing to offer a lower price in their spot markets because spot instance are *damaged goods* (Deneckere and Preston McAfee 1996); their functionality has been impaired, in this case by the risk of termination (Abhishek et al. 2012). A supercloud effectively undoes this damage (by making spot instances effectively as reliable as standard instances), which is bad for this pricing strategy.

Research Questions How should cloud providers change their damaged good strategy? Should they drop it entirely or increase the damage, e.g. by terminating instances without warning.

2.5 Stochastic Demands

A Supercloud can run in the spot market because it is evict-tolerant: Given warning it can avoid termination and be evicted to run elsewhere at minimal cost. While evict-tolerance is bad for a damaged goods strategy by cloud providers, it can be good for them in another context. Demand for instances in a datacenter varies stochastically. Thus, cloud providers must overprovision so they can handle peak periods. While in theory overprovisioned capacity could be sold in a spot market, anecdotally it is not sufficient for providers to recoup costs in this fashion since excess capacity is expensive. Instead, the existence of evict-tolerant instances would allow providers to reduce the amount of excess capacity by instead dealing with peaks by forcing the evict-tolerant instances to move to another datacenter which is not having a peak at the same time (perhaps even another provider).

Effect on Users Users can seek out providers willing to offer them better deals in exchange for designing their applications in such a way that they are evict-tolerant. While doing so may not be suitable for all applications, this is a much more palatable task than the current difficulty of dealing with potential instance termination.

Effect on Providers Since this saves them money, cloud providers may be willing to offer a discount for evict-tolerant instances. Note that, unlike with the spot market where the migration is within a datacenter, such cross-datacenter migration is more expensive and not all applications face the same cost, so this is a case where a damaged goods strategy could be effective.

Research Questions What is the best way for cloud providers to take advantage of the fact that some instances are evict-tolerant?

3. Discussion

We have examined five ways the Supercloud can save money for customers, and how each would affect public cloud providers. Two of them, pricing power and spot markets, are good for the Supercloud user but may be undesirable for cloud providers. As a result, it is not clear how a cloud provider may respond to these pressures given the fluidity of Supercloud users. They could, for instance, co-opt aspects of Supercloud's functionality. Two others, increased utilization and geographic mobility, have mixed effects on providers, but on balance may be beneficial for cloud providers as well as users. The fifth, the ability of evict-tolerant instances to help cloud providers better handle stochastic demands, potentially represents a strong positive from their perspective, and represents an application where cloud providers may see benefits from cooperating with a Supercloud.

Availability

The Supercloud project website is located at <http://supercloud.cs.cornell.edu>, and the code is publicly available.

Acknowledgments

The authors are supported in part by AFOSR grants FA2386-12-1-3008, F9550-06-0019, by the AFOSR MURI Science of Cyber Security: Modeling, Composition, and Measurement as AFOSR grant FA9550-11-1-0137, by NSF grants CNS-1601879, 0430161, 0964409, 1040689, 1047540, 1053757, 1151268, 1422544, 1518779, 1561209, and CCF-0424422 (TRUST), by ONR grants N00014-01-1-0968 and N00014-09-1-0652, by NIST grant 60NANB15D327, by DARPA grants FA8750-10-2-0238, FA8750-11-2-0256, and D11AP00266, by MDCN/iAd grant 54083, and by grants from Microsoft Corporation, Infosys, Google, Facebook Inc., and Amazon.com.

References

- Announcing Reduced Pricing on Storage. <https://azure.microsoft.com/en-gb/blog/storage-price-match/>, a.
- Announcing Infrastructure Services GA and New Price Commitment. <https://azure.microsoft.com/en-us/blog/the-power-of-and/>, b.
- V. Abhishek, I. A. Kash, and P. Key. Fixed and market pricing for cloud services. In *7th Workshop on the Economics of Networks, Systems, and Computation (NetEcon)*, pages 157–162, 2012. doi: 10.1109/INFCOMW.2012.6193479.
- J. Anselmi, D. Ardagna, J. Lui, A. Wierman, Y. Xu, and Z. Yang. The economics of the cloud: price competition and congestion. *ACM SIGecom Exchanges*, 13(1):58–63, 2014.
- Y. Bachrach, S. Ceppi, I. A. Kash, P. Key, and D. Kurokawa. Optimizing trade-offs among stakeholders in ad auctions. In *Fifteenth ACM Conference on Economics and Computation (EC)*, pages 75–92, 2014. ISBN 978-1-4503-2565-3. doi: 10.1145/2600057.2602879. URL <http://doi.acm.org/10.1145/2600057.2602879>.
- R. J. Deneckere and R. Preston McAfee. Damaged goods. *Journal of Economics & Management Strategy*, 5(2):149–174, 1996.
- Q. Jia, Z. Shen, W. Song, R. van Renesse, and H. Weatherspoon. Supercloud: Opportunities and challenges. *SIGOPS Oper. Syst. Rev.*, 49(1): 137–141, Jan. 2015. ISSN 0163-5980.
- Q. Jia, Z. Shen, W. Song, R. van Renesse, and H. Weatherspoon. Smart spot instances for the supercloud. In *CrossCloud*. ACM, Apr. 2016.
- C. Kilcioglu and J. M. Rao. Competition on price and quality in cloud computing. In *WWW'16*. IW3C2, 2016. To appear.