Towards free market cloud computing

Jeremy Singer, Abyd Adhami, Colin Perkins
School of Computing Science, University of Glasgow
Glasgow, UK
Email: jeremy.singer@glasgow.ac.uk, abyd@adhami.co.uk, csp@csperkins.org

Abstract—**Although cloud computing has made great inroads in terms of consumer adoption, we argue that the paradigm still has a long way to go to reach an economically efficient free market state. We highlight four key problems with the current cloud computing approach. These are (1) coarse granularity of service provision, (2) limited control of resources, (3) inadequate support for small providers and prosumers, and (4) vendor lockin via APIs. In this position paper, we briefly outline interesting work that may lead to a solution, in terms of moving towards a more market-based approach to cloud resource allocation.**

I. INTRODUCTION

Big data and ubiquitous computing systems are built upon the foundation of cloud computing, which is now a well-established concept [1]. Industry investment in cloud is worth tens of billions of dollars per annum [2], [3]. A recent enterprise survey [4] shows that 68% of companies currently host less than 20% of their application portfolios in the cloud, so there is still massive scope for increased cloud adoption.

In this paper, we argue that the computing industry has not yet realized the full potential of the cloud computing model for service provision and delivery. We outline four problems with cloud computing that prevent its full adoption in a sustainable society driven by big data. Ideally, a free market for cloud is the desirable scenario. However the problems outlined in Section II are impediments to a free market. Briefly, these are:

1) service provision is insufficiently fine-grained.
2) a restricted subset of resources is virtualized.
3) micro-cloud providers are unable to enter the market.
4) the range of cloud platforms is heterogeneous and broadly incompatible.

We generally focus on infrastructure-as-a-service (IaaS) cloud computing since this is where we have most experience. Nonetheless we expect that similar arguments will apply to other levels of utility computing provision. In Section III we sketch out some directions for future research in this area.

II. PARADIGM PROBLEMS

This section considers four issues with current cloud computing provision. We argue that these and similar problems are preventing the establishment and efficient operation of a free market economy for cloud services.

A. Coarse Granularity of Resources

In terms of Infrastructure as a Service (IaaS) provision, most providers offer predefined, quantized units of computing resource. For example, the lowest specification of an Amazon Web Services (AWS) instance is currently *t2.micro* which features a single virtual CPU and 1 GB RAM [5]. Although the CPU bursting concept allows some rewards for unused CPU cycles, the rewards can only be claimed in terms of subsequent processing, rather than refunds or cheaper pricing.

Another quantized unit is the minimum time period for which instances can be commissioned. AWS customers must rent machines by the hour. Google Compute Engine customers are charged for at least 10 minutes. Although there are some more light-weight services, e.g. the asynchronous event-based AWS Lambda platform, such innovations are not as flexible and are not yet commonplace.

This quantization of service is inefficient, in terms of market economics. Very often, consumers will pay for more provision than they actually require. Another disadvantage is that any kind of statistical analysis, modeling or forecasting for computing requirements is less straightforward with non-continuous parameters.

Finally, the different instance types hinder fair comparison. Often, it is not possible to make direct comparisons between different IaaS providers since they are offering products that are neither precisely equivalent nor directly comparable.

B. Limited Provision and Control of Resources

In addition to limited granularity of resources, providers generally offer only a limited set of resources that can be provisioned and controlled. Compute resources and memory can easily be provisioned, and some platforms offer the ability to provision GPU or storage resources, but options to provision the network access tend to be severely limited. That is, cloud providers tend to virtualize compute, memory, and storage resources, but offer no real access to virtualized network services.

This makes it difficult, for example, to run applications that are sensitive to network latency and/or jitter in cloud environments, since there is generally no way to request quality of service measures be employed.

Similarly, despite massive investment and research into improving data centre network protocols, and provision of software defined networking infrastructure to ease management and improve control over resources, it is not generally possible for clients of a cloud computing service to access these new features. Protocol changes and virtualized SDN resources are limited to cloud providers, and not made available to customers. We need visibility into the software-defined network, to allow configuration of network services, and to enable a market for network service differentiation and network function virtualisation in the cloud.

IEEE computer society

C. Exclusion of Small Providers

According to a recent industrial survey [6] the *big four* providers of compute infrastructure, (Amazon, Microsoft, Google, IBM) control over 50% of the cloud market.

In terms of physical requirements, there is a great deal of overhead in building and maintaining a datacenter [7]. Significant costs include physical infrastructure such as electricity and cooling, but also costs for managing compute, storage, and network resources, and security. Potentially unexpected costs include compliance with ISO certification standards for cloud security.

These significant barriers to entry prevent small providers from participating in the cloud service provision market. Initiatives such as the Open Compute Project www.opencompute.org help, by sharing data centre designs and best practices across providers and (somewhat) lowering the barrier to entry, and software platforms such as OpenStack www.openstack.org and OpenDaylight www.opendaylight.org ease deployment and manageability, but there is still a massive infrastructure overhead and deployment cost.

Meaningful innovation is occurring in the micro datacenter field, e.g. Raspberry Pi clusters [8] and clouds [9], [10], [11]. However it is unlikely that such new technology will be adopted by the market as it presently stands.

A further feature of a free market is the existence of *prosumers*. These are individual who, at different times, buy (i.e. they are consumers) and sell (i.e. they are producers) in the market. In terms of cloud computing, these prosumers are end-users who contribute compute cycles. There is a precedent for this in terms of distributed research projects like BOINC [12]. Generally, volunteer computing is more closely associated with *grid* rather than *cloud* technologies. However it may also be applicable to consumer clouds, given the scope of available resource [13]. There are clear parallels with the electricity utility market. Electricity prosumers may have solar panel installations that feed energy back into the grid. In terms of cloud computing, there is no practicable technique to harness this massively distributed small-scale prosumer behaviour at present.

In the long term, the entry of small providers and prosumers to the market will require some kind of arbitration or cloud brokerage service. While brokerages exist currently (e.g. Cloudaroo, RightScale) they are entirely focused on major service providers and do not support prosumers in any way. Grid broker services [14] for eScience applications are much more flexible and feature discovery-based protocols for service provision.

D. Incompatibility of Platform Utilities

Each of the major cloud service providers presents its own custom APIs and functionality for major application-level support. Typical features include key/value storage, database and outgoing email. There are few cloud industry standards, so all vendors offer their own, mutually incompatible, APIs. This problematic scenario quickly leads to vendor lockin for cloud users. Any kind of software investment built on a commercial IaaS platform will have vendor-specific assumptions and dependences. Ideally, a vendor-neutral cloud facade [15] would address this difficulty. Although some facade-style solutions are available (e.g. Apache libcloud [16] or DeltaCloud [17]) none is widely adopted by customers or supported by all cloud providers. In the same way as operating systems have the concept of POSIX compliance, we need something similar for IaaS providers. This would make it easy to migrate services dynamically from one provider to another, based on fluctuating demand, prices and other higher level non-functional concerns. Initiatives such as Nuxi CloudABI [18] have potential here, since they provide a reasonably familiar programming interface, cut-down to essential services and augmented by a capability-based security model that is well suited to cloud environments.

One might argue that the lowest common denominator across all IaaS providers, e.g. a bare-bones x86 Linux server, might be the common facade. Then cloud customers can build specific services on top of this standard base layer, e.g. database, key/value store, etc. Such services could be deployed in Docker containers, which are increasingly supported by major vendors. However, the administration overhead for all these services has a high cost. This means the cloud customer actually *loses* many of the benefits of utility computing in the first place (i.e. minimal setup and configuration).

III. RESEARCH DIRECTIONS

As cloud computing continues to develop and gain significance, the financial and marketing aspects become ever more important. At present, a handful of large service providers (i.e. Amazon, Microsoft, Google and IBM) hold a majority share of the market, each with their own set of branded products and proprietary ecosystem. We argue that there is greater potential benefit in opening up this marketplace.

By treating IaaS cloud as a *commodity* and potentially representing it by financial instruments, it can be traded on an exchange in similar ways to how oil is traded presently. The rationale here is that storage or computing resources are generic and hence organizations can exchange or trade these, competing on price or like-for-like services on a trading platform. This will help create more fluid supply and demand dynamics in the market and bring more competition in this space. Sharma et al. [19] present an initial formulation and simulation of this concept.

There are variations in how cloud services are priced and while there are several commercial cloud marketplaces, e.g Spot Instances and Reserved Instance Marketplace by Amazon AWS and SpotCloud by Virtustream (now part of EMC), they are affiliated and specific to each vendor respectively. This has led to some organizations moving into this space and creating an active exchange market for cloud services. One such exchange for computing capacity is the Cloud Exchange AG [20] recently launched through a joint venture between Deutsche Börse and Zimory. This aims to standardize the way cloud services are offered and measured on the market. It does this by enabling multiple cloud providers to offer comparable services on a single platform with the ability to select the cloud servers' location and jurisdiction.

One of the main challenges in trading utility computing resource like a commodity on a trading platform is the lack of industry standardization, as discussed in Section II. IaaS

providers each have different ways of processing workloads and apply different pricing structures. Our ultimate objective would be to be able to trade compute resource in similar fashion as energy or oil, which is naturally equivalent despite different suppliers. The Börse Cloud Exchange is a small step in the right direction, however further work is required to reach this goal. One possibility is that national governments or international organizations may seek to regulate the utility computing market, similarly to the energy market—to provide a competitive market by legal means. The European Union anti-trust ruling against Microsoft in 2004 is an analogous case.

The immediate aim, from a systems deployment perspective, is to enable dynamic scaling and allocation of cloud infrastructure, together with a smart agent-based system that determines where/when to purchase additional resource in the pay-as-you-go cloud domain. This framework could evolve into a trading platform which allows anyone to setup and start selling cloud resource. Further, an independent rating system or index would be required, to calculate a metric for reliability and performance of each cloud provider operating as a seller in the trading platform.

IV. RELATED WORK

Early research on *agoric systems* [21] argues that decentralized market mechanisms are an effective way to manage resources efficiently in large, complex distributed computing environments. An agoric system provides means to share knowledge and coordinate action efficiently. Both of these concepts are required for effective utility computing.

In the original presentation of agoric systems, the authors focus on specialized distributed computation activities, such as garbage collection across trust boundaries [22]. We have also developed simple economic models for memory management in virtual machines [23], [24]. However broadly similar principles should be applicable to the more general resource management issues in IaaS cloud provision.

Many of the concepts and technologies that underlie cloud computing have their origins in grid computing, as Barker et al. observe [25]. Cloud may be seen as the logical convergence of grid and peer-to-peer distributed computing [26]. Resource management is a key component in grids—Krauter et al. present a good taxonomic overview [27]. In their taxonomy, some grid resource management systems feature decentralized scheduling with predictive state estimation based on economic pricing models. This is the style of computational resource market we anticipate as being most useful for cloud. Buyya [28] presents a distributed computational economy for resource trading in grid systems, based on quality-of-service metrics.

Pal and Hui [29] construct a formal model based on queuing theory to describe a multi-provider cloud market for a single application type. They present an analysis of game theoretic models to show the existence of Nash equilibria, which can drive dynamic optimal resource allocation. We argue that this kind of model is essential, although we would broaden the market to multiple application types.

Jennings and Stadler [30] present a survey of resource management for cloud computing. They highlight five research challenges in this area, one of which involves 'Economic Behaviour and Pricing Strategies'. They anticipate the application of dynamic pricing to control demand, on the part of both providers and users.

Recent work involves reframing cloud resource allocation in an economic context. For instance, Sanchez et al. [31] describe the design of a brokerage service for IaaS cloud provision. Mashayekhy et al. [32] present an auction system, with a simulated evaluation. We expect to extend this work for a multi-provider cloud market with a concrete implementation.

V. CONCLUDING REMARKS

In this paper, we argue that resource management for cloud computing is still particularly primitive. Problems include the coarse granularity of IaaS provision, the limited control of resources, the exclusion of small service providers and the incompatibility of higher-level application services. Such difficulties impede the development of a commodities trading platform for IaaS resource. We outline the direction for our research, towards a free market for IaaS resource.

REFERENCES

[1] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica *et al.*, "A view of cloud computing," *Communications of the ACM*, vol. 53, no. 4, pp. 50–58, 2010.

[2] "Cloud computing spending to hit $32bn in 2015," http://www.v3.co.uk/v3-uk/news/2405174/cloud-computing-spending-to-hit-usd32bn-in-2015 (fetched 24 Aug 2015).

[3] "Public cloud computing to reach nearly $70 billion in 2015 worldwide," https://www.idc.com/getdoc.jsp?containerId=prUS25797415 (fetched 24 Aug 2015).

[4] "Cloud computing trends: 2015 state of the cloud survey," http://www.rightscale.com/blog/cloud-industry-insights/cloud-computing-trends-2015-state-cloud-survey (fetched 15 Sep 2015).

[5] "Amazon EC2 T2 instances," https://aws.amazon.com/ec2/instance-types/t2/ (fetched 15 Sep 2015).

[6] "Big four IaaS providers now own half the market," http://www.eweek.com/cloud/big-four-iaas-providers-now-own-half-the-market.html (fetched 24 Aug 2015).

[7] A. Greenberg, J. Hamilton, D. A. Maltz, and P. Patel, "The cost of a cloud: research problems in data center networks," *ACM SIGCOMM Computer Communication Review*, vol. 39, no. 1, pp. 68–73, 2008.

[8] S. J. Cox, J. T. Cox, R. P. Boardman, S. J. Johnston, M. Scott, and N. S. O'Brien, "Iridis-pi: a low-cost, compact demonstration cluster," *Cluster Computing*, vol. 17, no. 2, pp. 349–358, 2014.

[9] F. P. Tso, D. R. White, S. Jouet, J. Singer, and D. P. Pezaros, "The glasgow raspberry pi cloud: A scale model for cloud computing infrastructures," in *Distributed Computing Systems Workshops (ICDCSW), 2013 IEEE 33rd International Conference on*, 2013, pp. 108–112.

[10] P. Abrahamsson, S. Helmer, N. Phaphoom, L. Nicolodi, N. Preda, L. Miori, M. Angriman, J. Rikkila, X. Wang, K. Hamily, and S. Bugoloni, "Affordable and energy-efficient cloud computing clusters: The bolzano raspberry pi cloud cluster experiment," in *Cloud Computing Technology and Science (CloudCom), 2013 IEEE 5th International Conference on*, vol. 2, Dec 2013, pp. 170–175.

[11] V. Valancius, N. Laoutaris, and L. Massoulié, "Greening the Internet with nano data centers," in *Proceedings of the International Conference On Emerging Networking Experiments And Technologies (CoNEXT)*. Rome, Italy: ACM, December 2009.

[12] D. P. Anderson, "Boinc: A system for public-resource computing and storage," in *Grid Computing, 2004. Proceedings. Fifth IEEE/ACM International Workshop on*, 2004, pp. 4–10.

[13] D. P. Anderson and G. Fedak, "The computational and storage potential of volunteer computing," in *Cluster Computing and the Grid, 2006. CCGRID 06. Sixth IEEE International Symposium on*, vol. 1, 2006, pp. 73–80.

[14] "Grid service broker: A grid scheduler for computational and data grids," http://www.cloudbus.org/broker/ (fetched 24 Aug 2015).

[15] B. Di Martino, G. Cretella, and A. Esposito, "Cross-platform cloud APIs," in *Cloud Portability and Interoperability*. Springer, 2015, pp. 45–57.

[16] "Apache libcloud project," https://libcloud.apache.org (fetched 24 Aug 2015).

[17] "Apache deltacloud project," https://deltacloud.apache.org (fetched 24 Aug 2015).

[18] "Cloudabi: Cloud computing meets fine-grained capabilities," https://www.bsdcan.org/2015/schedule/track/Security/524.en.html (fetched 15 Sep 2015).

[19] B. Sharma, R. K. Thulasiram, P. Thulasiraman, S. K. Garg, and R. Buyya, "Pricing cloud compute commodities: A novel financial economic model," in *Proceedings of the 2012 12th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing*, 2012, pp. 451–457.

[20] "Deutsche Borse Cloud Exchange AG," http://cloud.exchange (fetched 24 Aug 2015).

[21] M. S. Miller and K. E. Drexler, "Markets and computation: Agoric open systems," in *The Ecology of Computation*, B. Huberman, Ed. Elsevier, 1988, available at http://e-drexler.com/d/09/00/AgoricsPapers/agoricpapers.html.

[22] ——, "Incentive engineering: for computational resource management," in *The Ecology of Computation*, B. Huberman, Ed. Elsevier, 1988, available at http://e-drexler.com/d/09/00/AgoricsPapers/agoricpapers.html.

[23] J. Singer, R. E. Jones, G. Brown, and M. Luján, "The economics of garbage collection," in *Proc. International Symposium on Memory Management*, 2010, pp. 103–112.

[24] C. Cameron, J. Singer, and D. Vengerov, "The judgment of Forseti: Economic utility for dynamic heap sizing of multiple runtimes," in *Proc. International Symposium on Memory Management*, 2015.

[25] A. Barker, B. Varghese, J. S. Ward, and I. Sommerville, "Academic cloud computing research: Five pitfalls and five opportunities," in *Proceedings of the 6th USENIX Conference on Hot Topics in Cloud Computing*, 2014, pp. 2–2.

[26] I. Foster and A. Iamnitchi, "On death, taxes, and the convergence of peer-to-peer and grid computing," in *Peer-to-Peer Systems II*. Springer, 2003, pp. 118–128.

[27] K. Krauter, R. Buyya, and M. Maheswaran, "A taxonomy and survey of grid resource management systems for distributed computing," *Software—Practice and Experience*, vol. 32, no. 2, pp. 135–64, 2002.

[28] R. Buyya, "Economic-based distributed resource management and scheduling for grid computing," 2002, available as arXiv preprint cs/0204048.

[29] R. Pal and P. Hui, "Economic models for cloud service markets: Pricing and capacity planning," *Theoretical Computer Science*, vol. 496, pp. 113–124, 2013.

[30] B. Jennings and R. Stadler, "Resource management in clouds: Survey and research challenges," *Journal of Network and Systems Management*, vol. 23, no. 3, pp. 567–619, 2015.

[31] F. Sanchez, S. Al Zahr, M. Gagnaire, J. Laisne, and I. Marshall, "Compatibleone: Bringing cloud as a commodity," in *Cloud Engineering, IEEE International Conference on*, 2014, pp. 397–402.

[32] L. Mashayekhy, M. M. Nejad, and D. Grosu, "A two-sided market mechanism for trading big data computing commodities," in *Big Data, International Conference on*, 2014, pp. 153–158.