Mapping Cross-Cloud Systems: Challenges and Opportunities

Yehia Elkhatib
School of Computing and Communications, Lancaster University, UK

Abstract
Recent years have seen significant growth in the cloud computing market, both in terms of provider competition (including private offerings) and customer adoption. However, the cloud computing world still lacks adopted standard programming interfaces, which has a knock-on effect on the costs associated with interoperability and severely limits the flexibility and portability of applications and virtual infrastructures. This has brought about an increasing number of cross-cloud architectures, i.e., systems that span across cloud provisioning boundaries. This paper condenses discussions from the CrossCloud event series to outline the types of cross-cloud systems and their associated design decisions, and laments challenges and opportunities they create.

1 Introduction
Cloud applications are developed against a remote API that is independently managed by a third party, the cloud service provider (CSP). Instigated by changes, such as pricing, porting an application from consuming one set of API endpoints to another often requires a fair degree of re-engineering especially considering that even syntactically similar APIs could digress semantically [24]. As such, the increasing realisation of the inevitability of cross-cloud computing [30][9][10] led to various proposed solutions. As expected with such a nascent field, there is a certain degree of confusion arising from the use of non-convergent terminology: hybrid clouds, multi-clouds, meta-cloud, federated clouds, etc. The first contribution of this paper, thus, is to offer a coherent understanding of cross-cloud computing [4]. The second contribution is a classification [5] based on the terminology witnessed to date in this field along with prominent efforts of each, describing their modus operandi and commenting on their suitability and limitations, and how they relate to the responsibility of different stakeholders.

2 Why cross cloud boundaries?
A cross-cloud application is one that consumes more than one cloud API under a single version of the application. Let’s consider a few examples drawn from real scenarios where developers are faced with the option to work with different APIs, i.e., to cross cloud boundaries.

- Alan, an online service provider, finds that his user base is more fleeting than he planned for: web analytics indicates that a large proportion of users are accessing services through mobile devices and only for a few minutes (as opposed to hours as Alan originally envisioned). Alan decides to change how he manages his service infrastructure using ephemeral virtual machines (VMs) as opposed to dedicated long-life ones. He, thus, changes his business plan to employ a different CSP that charges by the minute rather than the hour, saving him hundreds of dollars each month in operational expenses.

- A company is consolidating some of its internal teams and, accordingly, their respective services will be unified into a single platform. Bella, the company’s Chief Information Officer (CIO), is in charge of this task. Her objective is to keep all internal services operational and as as frictionless to use during and after the transition. Bella finds that the teams to be consolidated have been using different public and private cloud infrastructures for various operations deep within their structure. This necessitates major changes to the underlying logic that handles task automation, service provisioning, resource management, etc.
• An online gaming startup Casus is rapidly expanding its user base. The cloud allows Casus to consume an increasing amount of resources as and when required, which is extremely advantageous. However, the cloud does not necessarily aid in providing an optimized service to users who are not relatively close to any cloud datacenters, such as those in the Arabian Gulf region, western Africa, or central Asia. In order to cater to such users, Casus has to use innovative techniques to maintain high quality of experience (QoE). One such technique is to expand the housing of logic and data beyond any one CSP, but instead to be able to relocate on demand to local CSPs whilst maintaining service operation across the different infrastructure substrata.

A common thread to these scenarios is change to the predetermined plan relating to service provisioning, use, or management. Different parts of the application (virtualized infrastructure manager, load balancer, etc.) would need to be changed to call different APIs. Change is, of course, part of business. Hence, the need for cross-cloud systems naturally grows greater as industries and societies increasingly use the cloud. Such change, however, entails fundamental changes to the communication behaviour to accommodate different semantics, charging models, and SLA terms. This is the core cross-cloud challenge.

Another commonality is the need to be free from long-term commitment. Many consumers choose the cloud for agility and elasticity. In the past few years, this was restricted to the boundaries of a single CSP but currently the trend is to transcend different CSPs. A recent survey [9] discovered that the “ability to move data from one service to another” ranked very highly as a concern raised by private sector SMEs as well as large organisations that use the cloud.

As such, a number of works in academia and industry have attempted to tackle this challenge using different strategies. Before attempting to categorize these works, it is perhaps important to point out the obvious: This is not a thesis for a universally uniform provisioning system. First, such “uber cloud” is unrealistic given the commercial nature of the market. Second, we believe it to be healthy to have a diverse cloud market where each provider brings a unique mix of specialized services that caters to a certain niche of the market.

3 Cross-cloud Dictionary

We identify four types of cross-cloud systems (Table 1) based on the prevalence in the literature and from experience chairing the CrossCloud1 workshop series. We aim to provide some clarity about the terminology, and to use this moving forwards when discussing challenges.

3.1 Hybrid Clouds

Hybrid clouds are inherently made up of dissimilar parts [25]. The onus here is on the system developer to amalgamate these sub-clouds for the purposes of their application. This involves some logic to determine which sub-cloud is to be employed and when. Such policy is coupled to the rest of the application to some degree, and could be implemented as a separate module within the application or as a self-standing proxy between the application and the underlying clouds. In any case, the policy enforcing element of the application will eventually interface with a finite set of cloud APIs. Such hardwiring is costly if said APIs are altered after application development. Another hidden cost seldom discussed emerges from the forward design incurred to predetermine the exact sub-clouds to be employed. This is a crucial step to minimize the chance of changing the set of sub-clouds after application deployment.

Although hybrid cloud architectures are abundant, their highly customized nature does not necessarily lend them to reuse by others.

3.2 Multi-Clouds

Like hybrid clouds, multi-cloud systems build applications by employing different autonomous systems. However, in this model management is achieved via abstraction which relieves some of the cross-cloud responsibility from the end system developer. It also introduces a level of portability, a shortcoming of hybrid systems.

There are different forms of multi-clouds such as common programming models and API translation libraries. In either case, agreement on a least common denominator API between the different cloud systems supported by such abstraction environments does mean loss of specialized services. Any need outside the unified abstract API (e.g., billing) is not supported and needs to be addressed individually by the developer. This obviously detracts from the value of such solutions.

A notable example is the mOSAIC programming model [29]. Most of the traction multi-cloud engineering has had, however, came through the use of open source API translation libraries such as Libcloud, a Python toolkit that provides abstraction to 50+ compute and data service offerings. Alternatives include jClouds (Java, Clojure), Fog (Ruby), DeltaCloud (Ruby), pkgcloud (node.js), and elibcloud (an Erlang wrapper for Libcloud).

3.3 Meta-Clouds

Meta- or inter-clouds shift the responsibility even further away from the application developer, offering both abstraction and delegation. Meta-clouds are typically de-
Table 1: The different types of cross-cloud systems.

<table>
<thead>
<tr>
<th>Similarity of sub-clouds</th>
<th>Hybrid Cloud</th>
<th>Multi-Cloud</th>
<th>Meta-Cloud</th>
<th>Federated Cloud</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstraction</td>
<td>None</td>
<td>Some</td>
<td>High</td>
<td>Very high</td>
</tr>
<tr>
<td>Provisioning Means</td>
<td>Bespoke logic</td>
<td>Least common denominator API via a common programming model or a translation library</td>
<td>Delegating infrastructure instrumentation</td>
<td>Common API</td>
</tr>
<tr>
<td>Responsible Party</td>
<td>System developer</td>
<td>CPM or library developer</td>
<td>Third party brokers</td>
<td>Component cloud service providers</td>
</tr>
<tr>
<td>Examples</td>
<td>(domain specific; e.g., [14])</td>
<td>Libcloud, jClouds, Fog, MODAClouds IDE</td>
<td>OPTIMIS, Jamcracker, Dell Cloud Marketplace</td>
<td>CIMI, OCCI, TOSCA, CDMI</td>
</tr>
</tbody>
</table>

employed through third party brokers that provide loosely-coupled interaction as a managed service [33, 35]. Such brokers handle the discovery of suitable resources and subsequent life cycle management.

Some of these brokers evolved from meta-schedulers in grid computing (e.g., [2]) where infrastructure instrumentation is out-sourced to a centralized scheduler. An example is the work by Tordsson et al. [36] on a broker that optimizes placement and management of VMs across infrastructures. Other efforts approach brokerage by being more engrained into the application development process. The Optimis Toolkit prototype [15] constructs services as a configuration of core elements using a programming model and runtime. Each element is attached to a set of functional and non-functional requirements that are to be honoured by a central deployment manager that negotiates suitable resources.

3.4 Cloud Federations

A common condition in all of the above models is the lack of agreement amongst the sub-clouds. Federated clouds, on the other hand, achieve distribution through prior agreement that could be manifested as compliance to standard interfaces and data formats.

To highlight one example from academia, Celesti et al. [8] envisioned that CSPs will federate horizontally to gain market share and to further improve their capabilities by harnessing economies of scale. The Cross-Cloud Federation Manager (CCFM) is the solution they put forward to bridge the gaps between different providers, and is based on three stages: discovery, match-making and authentication. The solution is to be adopted by an CSP as a middleware to enable federation with other CSPs. A similar model came out of the RESERVOIR research initiative [31].

Standardisation efforts have aimed at creating a unified interface for working with resources and services across vendors. Examples include: the Cloud Infrastructure Management Interface (CIMI) [11], the Open Cloud Computing Interface (OCCI) [26], the Topology and Orchestration Specification for Cloud Applications (TOSCA) [5], and the Cloud Data Management Interface (CDMI) [11]. Other efforts have attempted to develop minimalistic APIs to support the most common interactions; e.g., [12, 20]. Despite the presence of cloud standardisation groups, their efforts are yet to pay off in terms of wide market adoption [28]. The major CSPs probably see standards as giving customers the opportunity to effortlessly switch to competitors.

4 Challenges

We have elucidated how building cross-cloud systems does not come without overheads. Until the rise of cost-effective brokerage services that a developer can offload the cross-cloud burden to, there will remain a significant development effort required to enable applications to seamlessly cross provisioning cloud boundaries. Additional costs and aftermaths are discussed here.

4.1 Developer Preferences

The choice between cross-cloud solutions is a trade off in which the application at hand plays an obviously big role. The development trade off boils down to the amount of flexibility required and the extent of time available. Hybrid clouds provide the most flexibility as the developer can build a custom tailored infrastructure using native APIs from any provider as and when required. On the other hand, this is a relatively significant development effort required to enable applications to seamlessly cross provisioning cloud boundaries. Additional costs and aftermaths are discussed here.
something an active open source community could help with. Meta-clouds provide even more abstraction to the developer at the cost of less control, but these solutions are currently lacking in the market as will be discussed in \[4.3\]. Finally, infrastructure management through cloud federation is easiest for a developer, but is yet to be widely adopted by the major CSPs.

4.2 Value-added Services

By definition, building a cross-cloud application using any of the solutions discussed in \[3\] results in a system that is somewhat unhinged from the underlying infrastructure. To the developer, this incurs additional overhead in terms of manually managing infrastructure resources. Using high level added-value CSP services (such as AWS ElastiCache or Google DataFlow) becomes less of an option as they tether the system to a certain ecosystem. It is exactly this sort of binding that unintentionally grows with time, building a gravity towards a certain CSP above others.

This cost is the reason why to many building cross-cloud applications might be great as a concept but is realistically unlikely. In a sense, it dissolves the golden promise of the cloud computing paradigm as proffered by CSPs: “don’t worry about the low-level system, we’ll take care of it; you just focus on your business”.

4.3 Mind the Broker Gap

Another unsavoury reality is the state of the brokerage market. Despite the huge appeal for customers and the seemingly incessant forecasts about its potential (cf. \[3\]), the market is in fact hopelessly slim. This could be attributed to a number of reasons, not least of which is entrusting a seemingly opaque third party (broker) to have huge influence about how the customer deals with a larger and even more opaque body (CSP).

However, we find that this is not the core problem as evident by abundance of brokers in other markets that portray similar settings (e.g., airline ticketing). The key disabler here is expertise. A business that possesses the in-house expertise to manage different cloud infrastructures in a hybrid cloud setup (e.g., Netflix) does not need to spend additionally on brokerage provided by a third party. In fact, an in-house built system would naturally be much more informed, customized and flexible.

This argument then suggests that companies that do not have such expertise internally (e.g., young startups) are perhaps the prime consumers of cloud brokerage services. Alas, such SMEs are primarily focusing on market definition and horizontal expansion, and thus might not have cross-cloud computing as a priority. Consequently, the brokerage business model is difficult \[7\]. Additional added value is required for the brokers to exist \[4\]. We discuss one such added value in \[5.1\].

5 Opportunities

In light of the highlighted trends and challenges, we now discuss some of the main opportunities and associated open questions. We exclude the subject of standard adoption as it has been well covered by others (cf. \[34\]).

5.1 Strategic Decision Making

Customers moving between CSPs (and indeed those entering the market) need assistance in choosing the right services for their requirements and constraints. This creates an opportunity for non-partisan consulting services that provide automated recommendations about which cloud services to use, and when and where to migrate.

Recent advancements in machine learning provide great opportunities here \[32\]. Such sophisticated decision making has been until very recently hindered by the cost of benchmarking data. However, cloud performance metrics are now attainable from sources like NewRelic\(^3\), ThousandEyes\(^4\), and CloudHarmony\(^5\). The evidence-based consulting frameworks envisaged here would enable additional services such as arbitrage and life cycle management, improving the brokerage market prospects.

- How to harness performance metrics to gain knowledge in different domains? \(i.e.\) between different CSPs, between different customer applications.
- What is the best way of quantifying the cost of migrating between CSPs? \(e.g.,\) \[19\].
- How to provide strategic consultation for a large number of customers while avoiding/minimizing thrashing and maximizing overall utility?

5.2 Crowd-sourced Cloud Knowledge

Companies such as NewRelic and ThousandEyes (mentioned in the previous subsection) make a business out of finding user-reported cloud performance metrics. CloudHarmony perturbed the market by openly disseminating benchmarking data they collect. In a cross-cloud capable world, such information is transformative as it allows users to vote “with their feet”, switching providers whenever they do not receive the service they are paying for.

Furthermore, cloud metrics are becoming even more accessible and comparable using tools such as Google’s open source PerfKit\(^6\). This allows customers to quantify and compare experiences about CSP usage, potentially creating a crowd-sourced knowledge base. Further, Microsoft is leading the way in obtaining billing information via an API\(^7\).
The next logical step is to provide means of verifying whether SLAs are breached or not. Indeed, researchers’ attention has started to turn towards means of programmatically verifying SLAs [21]. Coupled with crowdsourced knowledge, this would empower customers and help them make informed choices about competing services. It also has the potential to change the market by putting pressure on CSPs and increase competitiveness.

- How can we accumulate knowledge of how users consume resources and the QoS they experience? Could this be matched with application profiles and usage context to provide a fine-grained performance insight?
- Could such community knowledge be used to verify SLAs in an automated manner?
- What would be the effect of such evidence-based culture on the distribution of market share between CSPs of different sizes? How would it affect their service offerings and pricing, if at all?

5.3 Edge Market Places

The need for cross-cloud computing towards the edge of the network is increasingly important as more resources are available nearer to the end users [37, 23, 16]. This is especially true for locations where users are geographically distant from data centers [13], e.g., emerging economies with rapidly growing consumer markets but with no close data centers. This puts such promising markets at a disadvantage. Major CSPs are racing to set up data centres in such areas (cf. [27]). However, such expansion is a highly expensive, long term investment. Moreover, it cannot reach all users due to the sheer speed of growth of certain markets. Hence, there is a need for deploying cloud applications towards the edge.

This has been discussed for some years under different names including ‘cloudlets’ and ‘offloading’. However, previous efforts have been primarily motivated by the convergence of mobile and cloud computing and as such were focused on housing particular applications (e.g., mobile backend services) with certain stakeholders, such as ISPs. What we discuss here is fundamentally different. Cross-cloud computing at the edge is growing from the bottom up where the application could practically be of any type or scale, and the stakeholders are largely the consumers themselves. In such ecosystem, there is room for forming ‘community clouds’ as well as for entrepreneurs to provide ‘mini-clouds’ to serve local markets. This, in essence, is the rise of fog computing: the diminishing reliance on cloud data centers and the growth of heterogeneous computing infrastructures nearer to the end users.

- What programming abstractions are needed in order to easily straddle edge infrastructures? e.g., [38, 6]
- How to minimize gravity to any edge platform and ease difficulties of moving between infrastructures?
- Could major CSPs enter the fog market by providing services outside of their mega-data centers?

5.4 Migration Vehicles

Undoubtedly, virtualization plays a fundamental role in cloud computing, enabling the dynamic provisioning of bespoke environments. However, VMs as packaged memory and disk state are too expensive for cross-cloud purposes, both in network and operation terms. Containers (e.g., Docker), rkt, and unikernels provide lightweight alternatives. These are much more appropriate as portable vehicles to (re)deploy applications. VMs provide immutable infrastructure which is cumbersome to modify and manage once the VMs are ‘baked’. In contrast, containers and unikernels have good exposure to the developer and are thus much more easily controllable. They are well integrated with different infrastructure management tools like Puppet and Vagrant. In addition, containers and unikernels are much more efficient in sharing resources at the OS level compared to hardware virtualization by hypervisors. In effect, this opens up migration possibilities with more hosting destinations.

- Which of these technologies is the best vehicle for migrating different workloads?
- How to use containers to migrate operational workloads across infrastructures with ease whilst maintaining low network overhead? Are stateful containers (e.g., [17]) a viable solution?
- What does life cycle management of containers and unikernels across disparate hosting infrastructures look like?

6 Concluding Remarks

We gave an overview of cross-cloud computing, putting forward a dictionary to formalize the differences between solutions proposed in the literature. We discussed challenges and opportunities unravelled by the emergence of cross-cloud solutions. Cloud computing began and remains an industry-driven domain, but there is plenty of room for work done by system engineers not necessarily working with the big cloud players.

7 Acknowledgments

The author thanks participants and co-organizers of the CrossCloud event series, and the anonymous reviewers of this paper. Organizing CrossCloud events was partly supported by the Faculty of Science and Technology at Lancaster University.
References


Notes

2 “One Interface To Rule Them All”, as described on the Libcloud webpage: https://libcloud.apache.org/
3 http://newrelic.com/
4 https://thousandeyes.com/
5 https://cloudharmony.com/
6 https://github.com/GoogleCloudPlatform/PerfKitBenchmarker/
7 https://msdn.microsoft.com/library/azure/leab323-54db-423d-916f-190de96c8a3c
8 https://docker.com/
9 https://coreos.com/rkt