Ville Pekkala

VENDOR LOCK-IN PROBLEM IN CLOUD STORAGE

JYVÄSKYLÄN YLIOPISTO
TIETOJENKÄSITTELYTIETEIDEN LAITOS
2015

TIIVISTELMÄ

Pekkala, Ville
Vendor lock-in problem in cloud storage
Jyväskylä: Jyväskylän yliopisto, 2015, 22s.
Tietojärjestelmätiede, kandidaatintutkielma
Ohjaaja(t): Luoma, Eetu

Pilvitallennuspalvelut tuovat datan tallentamiseen ennennäkemättömiä etuja ja niiden hyödyntäminen on, sekä yksittäisillä kuluttajilla, että yrityksillä, yleistymässä. Teknologiassa on kuitenkin vielä ongelmia ilman ratkaisua. Turvallisuus-, sekä yksityisyysseikat ovat tunnetuimpia, ja niitä on tutkittu laajalti. Vähemmälle huomiolle on jäänyt, kuinka vaikeaa on palveluntarjoajan vaihtaminen. Palveluntarjoajan vaihtaminen aiheuttaa aina kustannuksia asiakkaalle. Tilannetta, jossa vaihtokustannukset ovat niin suuret, että ne estävät palveluntarjoajan vaihtamisen, kuvataan termillä "vendor lock-in" (toimittajaloukku). Tämän kandidaatintutkielman tavoitteena on kirjallisuuskatsauksen perusteella selvittää, mitkä aiheuttavat vaihtokustannuksia pilvitallennuksessa, mitä riskejä toimittajaloukku aiheuttaa sekä yksittäisille asiakkaille, että yrityksille, sekä kuinka toimittajaloukkua voi välttää.

Suurin osa vaihtokustannuksista tuli kirjallisuuskatsauksen perusteella datan siirtämisestä uudelle palveluntarjoajalle. Eriävät standardit estävät tallennetun datan helpon siirtämisen. Vaihtaessa palveluntarjoajaa, pilveen tallennettu data täytyy ensiksi ladata vanhalta palveluntarjoajalta ja sen jälkeen lähettää uuteen. Big Datan aikakaudella tämä voi olla hyvin kallis ja aikaa vievä operaatio. Muut kustannukset liittyivät uuden palvelun käytön opettelemiseen.

Asiakas on toimittajaloukkuun jäädessään luonnollisesti pilven toimittajan armoilla. Hinnannousujen tai palvelutason muutosten sietäminen ovat yleensä pienempi paha kalliiseen muutto-operaatioon verrattuna. Tämän lisäksi, jos data on keskitetty yhteen paikkaan, asiakkaalla ei ole pääsyä dataan toimittajan kärsiessä palvelukatkosta.

Toimittajaloukkua välttääkseen täytyy asiakkaan ottaa se huomioon jo pilvitallennuspalvelua valitessa. Yksi valintakriteereistä pilvitallennuspalvelulle pitäisi olla pilvestä poistumisen helppous.

Datan hajauttaminen useampaan paikkaan on luonnollisesti myös ratkaisu ongelmalle. Hajautetulle pilvitallennukselle on kehitetty useita menetelmiä. Ne hyödyntävät RAID:n kaltaisia menetelmiä jakamalla tiedoston osiin usealle pilvialustalle siten, että vain osa tiedostosta riittää sen uudelleenrakentamiseen.

Huolimatta keinoista, joilla minimoida toimittajaloukun riskejä, toimittajaloukku horjuttaa pilvitallennuksen hyödyllisyyttä sekä pysyy ongelmana, joka on syytä tiedostaa harkitessa datan tallentamisen ulkoistamista pilveen.

Asiasanat: toimittajaloukku, pilvitallennus, vaihtokustannus, hajautettu pilvitallennus, big data.

ABSTRACT

Pekkala, Ville
Vendor lock-in problem in cloud storage
Jyväskylä: University of Jyväskylä, 2015, 22 p.
Information Systems, Bachelor's Thesis
Supervisor(s): Luoma, Eetu

Cloud storage brings unprecedented benefits to data storage. This has been noted by both individual consumers as well as organizations as the utilization of cloud storage services is growing. Technology still has its' problems. Security and privacy aspects have lots of research done. Another important deterrent to the adoption of cloud storage, and one which has not received as much attention, is how hard it is to switch cloud storage provider. A situation, in which switching costs are so high that they block vendor switch, is called vendor lock-in. This bachelor's thesis means to investigate what are the causes for switching costs, what risks does vendor lock-in expose both individual consumer and organizations into and how one can avoid getting locked-in.

According to the literature review, biggest cause for switching cost comes from the laborious data migration process. Differing standards obstruct easy data migration. Instead, when switching service provider, the customer has to download data from the old cloud service provider and then upload it to the new cloud provider's servers. In the age of Big Data, this can be very expensive and time-consuming operation. Costs also originate from having to learn to use the new service.

Locked in customer is at the mercy of the cloud vendor. Tolerating price hikes or changes in service level is usually a better alternative than undertaking the costly migration operation. Vendor lock-in also causes customer to be more vulnerable to service outages as he has all his important data centralized in one place.

To avoid vendor lock-in, customer has to be aware of the issue when choosing a service provider. One of the criterion when choosing a vendor needs to be how easy it is to get the data out of the cloud.

Distributing the data to multiple locations has been identified as a solution to minimize vendor lock-in problem. Distributed cloud has multiple implementations. They utilize RAID-like technologies by splitting file over multiple different cloud storage providers' servers so that only a certain part of the file is needed for reconstruction.

Despite the ways to minimize the risks, vendor lock-in still remains a problem that has to be acknowledged by customer looking to outsource data storage to the cloud and one that slows down the adoption of these technologies.

Keywords: vendor lock-in, cloud storage, switching costs, distributed cloud storage, big data.

FIGURES

SISÄLLYS

1 INTRODUCTION

Advances in information technology and the diffusion of high-speed internet have made it possible that data can be stored to, and quickly retrieved from, third party owned cost-effective datacentres (Armbrust et al., 2009). This model where computing resources can be accessed ubiquitously, conveniently and on-demand via network access is called *Cloud computing* (Mell & Grance, 2011).

Cloud computing has several different models. This paper will focus on *cloud storage* aspect, a cloud computing model where cloud storage provider (CSP) rents customers storage space in its own servers and makes it accessible to the customer over the Internet (Janssen, 2015). Cloud storage customer can access and manipulate the data via a provided web-based interface (Wu, Ping, Ge, Wang, Fu, 2010). For example, Amazon S3 provides the client with an application programming interface (API), which enables number of commands to manage the data, such as copying or deleting it from the server (Amazon S3, 2015). Storage space is usually rented on a monthly basis for every gigabyte stored. Customer is also billed a bandwidth fee for accessing that data. (Abu-Libdeh, Princehouse & Weatherspoon, 2010) Some of the commonly used CSPs include Amazon S3, Google Storage, Rackspace, and GoGrid.

Outsourcing storage to a cloud vendor brings many appealing benefits to the data owner. For one, the data can be accessed from anywhere at any time (Armbrust et al., 2009). Secondly, the cloud user only has to pay for the amount of storage space he is using, and the cloud storage space scales rapidly based on demand. This also means that there is no up-front commitment or cost of managing hardware (Armbrust et al., 2009). Cloud storage is a convenient way to store large amounts of data without huge upfront commitments that are unavoidable when buying physical storage space.

In the age where high definition videos and pictures can be captured anywhere at any time with smartphones, the amount of data that consumer needs to store is getting massive. According to ICT consulting company Gartner's forecast (2012), one third of this data will be stored to cloud by 2016. Organizations are becoming aware of the benefits of cloud storage as well. According to a study done by EMC Corporation (2014), current trend shows rapid growth in utilization of cloud storage services for companies. Study shows a 100% of increase in the use of external cloud in three years, from 3% of total data saved to public cloud in 2012 to 6% in 2015.

As the use of cloud storage services is growing, the issues associated with them are becoming more important to acknowledge. Commonly cited literature on topic of cloud storage services usually seem focus on security and privacy aspects (Bowers, Juels & Oprea, 2009; Wang, Ren, Lou & Li, 2010). Another important issue in cloud storage, and the one this paper focuses on, is *vendor lock-in*.

Vendor lock-in is a term for a situation in which customer using the product or service cannot easily switch service provider without considerable *switching costs* (The Linux Information Project, 2006). According to Abu-Libdeh, et al. (2010), switching costs are largely caused by the migration of data from the cloud to another service provider. Along with being an inconvenience, moving data from one cloud server to another also causes severe financial losses. They say that "storage providers charge for both inbound and outbound bandwidth, [...] customer moving from one provider to another pays for bandwidth twice along with the actual cost of online storage".

Due to the increasing utilization of cloud storage by both individual customers and companies, vendor lock-in is becoming larger issue than before. The objective of the thesis is to provide a comprehensive presentation on the issue and the risks vendor lock-in exposes customers into. The thesis also looks into how to avoid the lock-in situation. This thesis will prove that vendor lock-in is a very important issue for the cloud storage customer, and that it needs to be acknowledged when making a decision to purchase said services.

The objective of the thesis is to answer the following questions:

- "What causes switching costs in cloud storage?"
- "What are the risks associated with vendor lock-in in cloud storage services?"
- "How can a cloud storage user avoid vendor lock-in?"

This paper answers these questions by doing a literature review of vendor lock-in and cloud storage. This review utilizes research papers found from reputable publications. To find these papers, it uses services such as IEEE Xplore and Google Scholar to find out what are the most cited papers of the subject. AIS Electronic Library is also used to find papers from the subject of computer science. For the literature review, this paper follows the guide for systematic literature review proposed by Okoli and Schabram (2010).

Typing "vendor lock-in + cloud storage" into Scholar yields 708 results. Results found using search terms "cloud storage risks" mostly focused on the data's security and privacy aspects in the cloud. The most cited paper on the subject is one by Abu-Libdeh et al. (2010), which is also used as the main source in this paper. Their paper provides important reasons for avoiding vendor lock-in and present a solution, distributed storage method that will be expanded upon in chapter 3. Distributed cloud method, which has been identified as a mean to avoid vendor lock-in (Toosi, Calheiros, & Buuya, 2014), has received a number of different proposals that tackle multiple problems in cloud storage. Other literature included in this literature review is found with the method described above and are somehow relevant to the research questions.

In the next chapter the thesis will look into what are the causes for switching costs in the cloud storage context. The following chapters will focus on the risks vendor lock-in brings and look into how vendors may propagate the lock-in, respectively. Chapter 3 will focus on ways to minimize the risk of vendor lock-Chapter 3.2 will also expand upon distributed cloud storage method, and how it can minimize the risks of vendor lock-in. The final chapter 4 will conclude the paper.

2 VENDOR LOCK-IN

Vendor lock-in is one of the most important issues in cloud storage that doesn't perhaps receive as much attention as security and privacy aspects. When customer is forced to use one service, he is vulnerable to number of risks. This chapter provides answer for the first research question, what factors cause switching costs in cloud storage; why does a customer become locked in? CSPs also have some policies that further encourage lock-in that are also examined in this chapter. This chapter also answers the second research question of what risks being locked in exposes customers to.

To avoid lock-in in cloud storage, it is very important to be aware of how customers become locked in, and what risks it causes to the customer. This chapter answers both of these questions and proves that vendor lock-in is definitely an issue that has to be taken to consideration.

2.1 Vendor lock-in in cloud storage

Vendor lock-in has been identified as a huge issue in cloud computing; Toosi et al. (2014) claim that it is one of the biggest problems in regards to cloud computing adoption. Neal Leavitt (2009), in his review about cloud computing issues from a company’s point of view, notes that even if the customer is dissatisfied with cloud provider, or the vendor goes out of business, the firm cannot "easily and inexpensively transfer service to another provider or bring it back in-house".

As it was mentioned earlier, vendor lock-in is a term from economics signifying the difficulty of switching from one service provider to another. Outsourcing data management to a CSP brings attracting data management cost reductions. However, every time the customer stores data to the cloud, the more dependent he becomes on that cloud vendor. Eventually it might lead to a situation in which the customer has so much data stored to the cloud that it effects the decision to change the CSP.

If the customer becomes unhappy with the service or if another provider started to offer storage services at a better price or better service level, question arises on how to get the stored data from cloud to the new service provider. There are no widely adopted standards that would let the customer move the data freely among different CSPs (Taneja, 2012). Indeed, in order to switch storage provider, cloud customer has to first download the data from the cloud and then upload it to the new CSP. This laborous migration process is the main cause for switching costs that will be expanded upon on the next chapter.

2.2 Cloud storage switching costs

In cloud storage, moving data from one provider to another brings huge switching costs. Abu-Libdeh, et al. (2010) note that because storage providers charge clients for both inbound and outbound bandwidth, the customer has to pay the bandwidth twice during the migration process. In addition, during the migration, the customer has to pay for storage capacity to both CSPs (Abu-Libdeh et al., 2010).

Besides having to pay to extract data from cloud, the process can be time consuming. Enterprise storage company Nasuni's (2014) data migration test revealed that the transfer of 12 TB of data from one cloud to another could take from few hours to almost a week, depending on the cloud's write capability. Obviously, not having access to data for days would be catastrophic for a company. Furthermore, according to an article by George Crump (2014), some service providers may also have some bandwidth limitations in place that further slowdown the migration process.

These factors together can cause huge financial switching costs to the cloud customer who has large amounts of data stored in the cloud. Companies may have data saved in the magnitude of petabytes to the cloud which is very inconvenient to extract. The increasing consumer utilization of cloud storage services makes this a significant issue for them as well.

Other CSP may offer cheaper storage services for the customer, but the costs of data migration prevent the switch. Abu-Libdeh, et al. (2010) provide a visual representation of the financial switching costs in Figure 1. It demonstrates a situation in which switching providers sounds reasonable by comparing their pricing schemes but the switching costs cause vendor switch to be very costly.

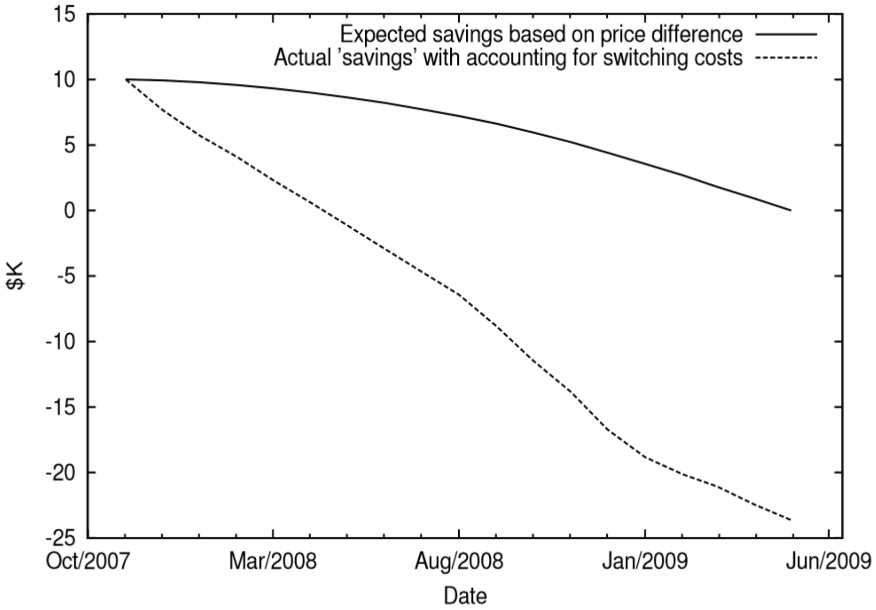

Figure 1: "Month-by-month switching benefit (non-RACs solution)" (Abu-Libdeh, et al. 2010).

According to paper about IT outsourcing by Whitten (2010), switching costs can also be psychological. Changing a service provider brings uncertainty to a company. Whitten notes that managers may hold certain expectations about the service provider, "but the the gap between expectations and knowledge represents a risk or cost of switching".

There are also no strictly enforced cloud computing API standards (Armbrust et al., 2009; Leavitt, 2009). This means that the differing APIs between different storage service providers can cause inconvenience from having to learn to use a new service (Whitten, 2009).

Silva, Rose, & Calinescu (2013), mention differences in semantics along with differing interfaces as a factor affecting cloud migration. According to them, cloud providers have differing descriptions of their services that confuse the customers. Differences in API and semantics might mean costly training programs to the organization staff operating the storage services.

Cloud storage vendor has the customer in a tight leash. Other vendor might offer better services at a more competitive price but the switching costs caused by the data migration and having to learn new service make the switch infeasible. In many cases the switching costs far outweigh the potential financial savings gained from switching the provider. However, customer has to ask if enduring the switching costs to change the situation is better alternative. As being locked in to one service provider brings many risks to the customer that are explored in the following chapter.

2.3 Vendor lock-in risks

Most terms of service let cloud provider increase change pricing at any time (Satzger, Hummer, Inzinger, Leitner, & Dustdar, 2013). As Abu-Libdeh et al. (2010) note, vendor lock-in obviously brings the vendor leverage over the clients. So much so that the clients are willing to reluctantly adhere to pricing increases in favour of the switching costs caused by vendor lock-in. As it was mentioned, CSPs usually charge for the storage space and for the bandwidth costs of accessing that data. With the massive amount of data that may be stored to the cloud, even a small increase in storage or bandwidth charge could bring massive costs to the customer. For example, say that a company has a hundred terabytes of data saved to the cloud. One cent of increase in the cost per gigabyte leads to 1000 dollar of increased cost per month.

Many of the CSPs advertise high data availability and reliability, making claims that cloud storage is more secure than physical saving forms (Google Cloud Storage, 2015; Amazon S3, 2015). However, as Abu-Libdeh et al. (2010) mention, although improbable, unforeseen events may cause service outages. Company that is locked-in to a single provider is dependent on the availability of the cloud. If CSP suffers from an outage, company has no access to their data for the duration of the outage, and as Satzger et al. (2013) put it, "could be in a standstill until it comes back online". Lock-in also prevents the customer from leaving if the CSP performs at an unacceptable service level.

As Andrew Coats points out in his article about cloud storage in the aftermath of Megaupload closing (2012), the customer is always responsible for the data he stores in the cloud. He points out that almost all of cloud storage terms of service remove cloud providers from all responsibilities in the case of data loss or provider going out of business. Same as with physical storage forms, having all the data in only one place exposes the customer to the risk of losing everything permanently.

Armbrust et al. (2009) mention The Linkup as an example. The service lost almost half of its customer's data and subsequently went out of business (Brodkin, 2008). Customers who did not have a back up to their data lost it irretrievably. The lucky ones who did not lose their data were forced to perform a hasty migration without having time to prepare.

Vendor lock-in takes power away from the customer, ties his hands and makes the customer be at the mercy of the vendor. Vendor can get away with performing poorly or price increases. Service outages to a customer who is dependent on one cloud storage service, is at a risk of having their whole business be in a standstill. Although generally dependable, the cloud may be affected by a series of improbable events that cause customers to lose some of their data completely. Customer is even affected by the risk of the cloud vendor going completely out of business. In this scenario, the customer has to perform a sudden and hasty data migration without any planning. Being too dependable on one service exposes customers to a number of serious risks.

2.4 Vendor lock-in catalysts

Vendor lock-in is obviously very beneficial to the cloud storage provider so CSPs naturally have certain policies that encourage vendor lock-in. Getting data to cloud is generally made much easier than getting it out of there (Abu-Libdeh et al., 2010). For example, Amazon S3 offers a free trial and all data uploads to the cloud are free (Amazon S3). According to Abu-Libdeh et al. (2010), some CSPs even offer to import data to the cloud via transfer services. However, when it is time to get the data out of the cloud, the CSP is not as helpful. Same services are not available for the outbound customers and transitioning off CSPs platform is not made as easy as it could be (Moyle, 2012). No wonder, why would they want to help their customers move to their competitors?

Some CSPs might even be actively hindering the transition process (Moyle, 2012). One of these ways is restraining customer's access to the data (Moyle, 2012; Pearson & Benameur, 2010). In a case depicted by Chua, Lim, Sia, & Soh (2008), a website was unhappy with the outsourced IT company. The outgoing company acted with hostility during the transition period, holding the client hostage by claiming ownership on data needed during the period. The client had no choice but to renew the contract for a short-term for the duration of the transition.

The pricing of cloud storage vendors typically starts high and lowers significantly as the amount of data in the cloud increases to encourage lock-in. For example, GoGrid's cloud storage starts at $0.12 per GB per month. However, for every gigabyte after 500 terabytes in the cloud, the price is $0.09, 25% cheaper (GoGrid, 2015). This makes distributing the data in small chunks among multiple cloud providers considerably more expensive.

Silva et al. (2013) quote Govindarajan & Lakshmanan (2010), stating that cloud standards are regularly proposed as a way to mitigate vendor lock-in and achieve portability and interoperability in cloud computing. However, Silva et al. quote Petcu (2011), continuing that there has not been widely adapted standards as the CSPs are concerned about the loss of customers that may come with standardization.

Abu-Libdeh et al. (2010) mention another aspect of vendor lock-in to consider. CSPs such as Google, Amazon and Apple have lots of products they are offering. They do not want their customers to only get locked in to their storage services, but to their whole ecosystem. iCloud is designed to work best with the rest of the Apple products, being accessible only with Apple ID. Abu-Libdeh et al. (2010) provide an example of Amazon EC2 that "can read from and write to Amazon S3 storage with low latency and no bandwidth charges". These convenient supplementary products makes using one cloud provider a habit and makes switching provider much more inconvenient.

Vendor lock-in exposes the customer to a number of risks that undermine the usefulness of cloud storage. For example, cloud vendor is free to change pricing or service level at any point. Cloud storage customer is also at risks of service outages and data loss. Switching provider is made expensive by costs in data migration and differing standards and those costs force the customer to continue using the service despite these risks.

One of the benefits of cloud storage is that there are no up-front commitments, which gives the cloud customer market mobility. Vendor lock-in eliminates this benefit, and makes the customer committed to a service when he is saving data there. It is something that a customer needs to be wary of, and the next chapter focuses on actions customer can take to avoid this issue.

3 AVOIDING VENDOR LOCK-IN

Vendor lock-in is arguably risky for the customer and a risk that has to be minimized when utilizing cloud storage services. The next step is to figure out ways how to avoid it. This chapter will answer the final research question; "How can cloud storage user avoid vendor lock-in".

This chapter will list precautions the customer can take when making the choice of CSP to reduce the risk of vendor lock-in. The chapter will also present the distributed cloud storage method that has been identified as a way to avoid vendor lock-in.

3.1 Precaution

When choosing a CSP, the customer has to be mindful of the risk of vendor lock-in. In his article, Taneja (2012) lists some steps to take to reduce the likelihood of vendor lock-in. According to him, customer has to determine how CSPs facilitate moving customer data out of their cloud storage repository and whether it can be done in a reasonable time frame. Taneja further adds that customer should find out whether the CSP supports data migration tools that make cloud-to-cloud transfers possible. Taneja also suggests that CSPs which have pledged to support cloud computing standards should be preferred.

Moyle (2012) mentions in his article about vendor lock-in that "customer has to be mindful of ways that service providers might try to propagate lock-in, or at least where they might be less helpful to transition – and testing the processes that support a clean transition is a solid strategy for avoiding cloud computing vendor lock-in down the road".

Two common points in these two articles seem to be that the customer should find out in advance how hard it is to get out of the cloud. This is reinforced in an article by Rabbetts (2013), which handles vendor lock-in based on personal experience. According to Rabbets, their company's data migration process took several weeks. Rabbetts says that they learned a valuable lesson, that they should ask more questions about the data, how it is stored and used and how to extract it. He says that customer should find out in advance if something goes wrong, how the moving of the service to another provider happens in practice.

The consensus of these three articles are that customer should find out how to get out of the cloud in case the relationship ends. Just as pricing and SLAs, how the CSP handles getting out of the cloud has to be taken into consideration when choosing the storage provider. As Opara-Martins, Sehandi and Tian (2014) put it, having an up-front plan of how to exit the contract is the golden rule of outsourcing.

3.2 The distributed cloud method

Having the data in multiple locations instead of with only one CSP sounds like a intuitive protection against the risks of vendor lock-in. Replication has been indeed identified as a solution to the vendor lock-in problem (Razavian, Khandi, & Yazdani, 2013). Spreading the data among multiple providers makes the customer resistant to price hikes and service outages (Abu-Libdeh et al., 2010). It also gives the customer mobility, guards him against the risks of data loss, or corruption of data at a single cloud provider (Slamanig & Hanser, 2012). One way to achieve this is to simply store an entire copy of a file at each provider. As Abu-Libdeh et al. (2010) note, that while this achieves the goals of redundancy and market mobility, it brings huge bandwidth and storage costs to the customer.

Another, more economical method, is to disperse suitably encoded data over a number of service providers (Abu-Libdeh et al., 2010; Slamanig & Hanser, 2012). This method disperses a single file over multiple vendors so that only a certain amount of the data is needed for file reconstruction. This applies the RAID-like technologies used with disks and file systems for decades to cloud storage (Abu-Libdeh et al., 2010). This utilization of *erasure coding* to cloud storage “reduces the risk of being influenced from single cloud provider outages and provide (at least to some degree) resistance to loss or corruption of data at cloud providers” (Slamanig & Hanser, 2012). This method disperses the file "over *n* vendors so that *k* shares of the file suffice for reconstruction" (Slamanig & Hanser, 2012). This way, if one vendor happened to lose the data, data from other vendors can be used to reconstruct the file. This so called distributed cloud storage method has been presented as a solution to the problem in the literature (Abu-Libdeh et al., 2010; Slamanig & Hanser, 2012; Schnjakin & Meinel, 2013).

Even though protecting the customer from the risks of vendor lock-in, this method seems expensive. As the cloud providers offer discounts for large amounts of data, dispersing the data in small chunks across multiple vendors instead of centralizing data storage to one vendor sounds financially unintuitive. According to Abu-Libdeh et al. (2010), striping the data across 9 providers has an added overhead of 11% of the original monthly cost. They continue that in addition to increasing operation and bandwidth costs, this method also inreases latency for client accessing the data.

However, as AbuLibdeh et at. (2010) argue, in the long term, this method can bring a lot of cost savings. The distributed cloud method is resistant to price hikes, as depicted in Figure 2 (Abu-Libdeh et al., 2010). If one CSP decided to increase costs, overall storage costs are not affected much. If a CSP with better service level emerged in the market, moving just part of the data instead of making an "all or nothing" decision, has minimal switching costs, as depicted in Figure 3 (Abu-Libdeh et al., 2010). Naturally, the higher amount of providers the data is distributed over, the more resistant this method is to the price hikes, switching costs and service outages. At the same time, the overhead costs grow, however.

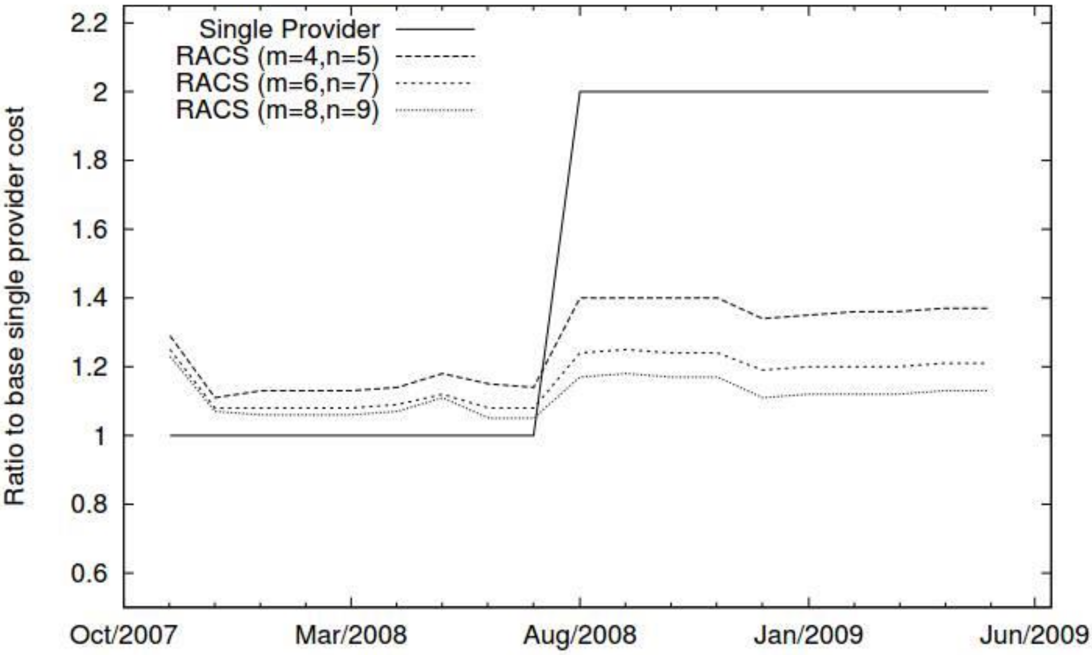

Figure 2: "Tolerating a vendor price hike" (Abu-Libdeh et al., 2010)

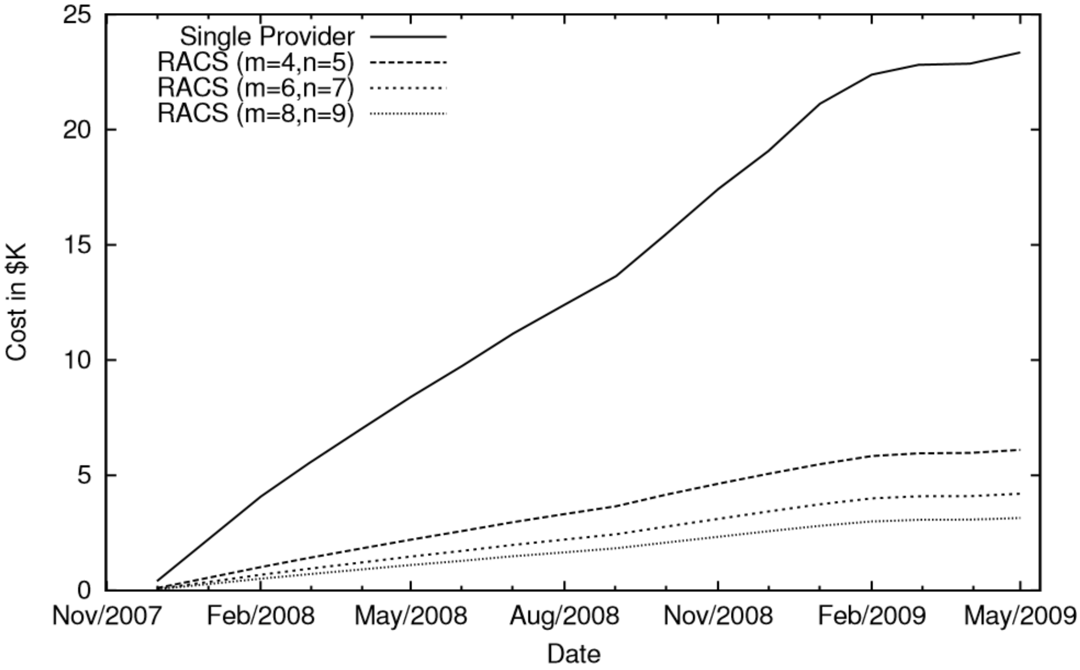

Figure 3: "Month-by-month swithing costs for various configurations" (Abu-Libdeh et al., 2010)

In addition to resisting financial risks of lock-in, one of the main advantages of distributed storage is that it is tolerant to data loss. As it was shown earlier, CSP is not responsible for losing its customer's data and although rare, it does happen. According to Abu-Libdeh et al. (2010), well executed erasure coding protects the customer from data loss of one or more of providers at a reasonable cost overhead.

Resistance to price changes, tolerating service outages, and tolerating data loss are the biggest advantages of distributed cloud storage. The main disadvantage of these methods are the increased costs. The more redundant the customer wants his data to be, the higher the storage, request and bandwidth costs get. Distributed cloud is also slightly slower as it is as fast as the slowest repository in it. Encoding and decoding striped data also causes latency. (Abu-Libdeh et al., 2010). Compared to the risks of vendor lock-in, however, the added costs and increased latency sound much better alternatives.

3.2.1 Different distributed cloud methods

In their research paper, Slamanig and Hanser (2012) look at a number of different distributed cloud methods and provide a brief overview of each method. Each method has different priorities and is suitable for different needs. RACS method, which has been discussed in earlier chapters, is meant to ensure the availability and durability of the data without excessive overheads (Abu-Libdeh et al., 2010). It does not try to improve, for example, confidentiality, integrity, or authenticity which is done by some of the other approaches to distributed cloud storage (Slamanig & Hanser, 2012).

Notable examples of distributed cloud methods include HAIL, which emphasizes availability and integrity by requiring the CSP to prove that a file is retrievable (Bowers, Juels, Oprea, 2009), Cloudproof, which focuses on proving cloud violations to third party, and Tahoe-LAFS which claims that the entire filesystem continues to function correctly even if some of the cloud providers fail (Slamanig & Hanser, 2012). Not present in the research paper is Cloud-RAID, which is very similar to RACS, but introduces a system to detect data corruption or confidentiality violations to improve security of the files (Schnjakin & Meinel, 2013). One thing that has to be considered when selecting a distributed cloud method, according to Slamanig and Hanser (2012), is how it handles multiple people accessing the data. Different distributed clouds have different ways to handle multiple users and some of them do not support multiple reader and writer environments.

This paper focuses on vendor lock-in issue and will not delve further into the designs of these methods. Any method of distributing data over many cloud vendors is a deterrent to vendor lock-in, even though each of the methods have different priorities. Cloud customer should choose the method that best suits his needs. Distributed cloud storage comes at the price of higher storage costs and increased latency in data access. The advantages it provides, however, are well worth the price.

Vendor lock-in is a huge problem in the field but this chapter has proved it can be avoided with careful planning and by having the data in multiple locations. There are plenty of methods that help the customer achieve this by using RAID-like erasure coding to disperse each file over multiple CSPs.

4 CONCLUSION

Cloud computing brings unpresented benefits to storing large amounts of data. Data stored to cloud can be accessed from anywhere at any time, the cloud is freely scalable and cheaper than conventional storage methods. It also has no upfront commitment costs deriving from hardware investments. The benefits of cloud storage are being utilized more by both individual customers as well as companies. However, cloud storage brings risks that need to be carefully inspected before making a decision to adopt the technology. The risk that this paper focused on was vendor lock-in which has been identified as a significant deterrent to the adoption of cloud storage together with privacy and security issues.

Vendor lock-in is the situation in which it is costly to switch service provider because of switching costs. The first research question asked what causes switching costs in cloud storage. The literature review identified data migration and differing standards as the causes for switching costs. The more data customer has in one cloud storage provider's servers, the harder it is to get that data out of there when it is time to switch service provider. Because of differing standards between cloud storage providers, switching provider involves downloading the data out from the old CSP and uploading it to the new CSP as well as having to learn to use the new service. It is always more expensive to get the data out of the cloud than it is to get into the cloud.

Vendor lock-in is a problem that keeps feeding itself. The switching costs keeps customers using the vendor which in turn further increases the switching costs. However, if better service is available, it may be preferable to endure the switching costs as being locked in to one vendor brings a lot of risks that have to be considered.

Being locked-in makes the customer dependent on a cloud storage provider and means that the customer is at the mercy of the vendor. This paper has examined numerous threats, i.e. financial costs, vendor losing data or going out of business, which the vendor lock-in brings to both company and individual customers. Customer has no choice but to suffer these risks as switching costs tie the customers' hands.

After identifying the causes for switching costs and the risks of vendor lock-in, the paper set to find ways to avoid vendor lock-in. The customer can also take pre-emptive action to minimize the risks of lock-in. When making a decision to outsource data storage to a CSP, customer needs to compare how easy it is to get the data out of the cloud along with comparing prices and service level agreement.

Replication has been identified as a way to avoid vendor lock-in. The customer should make sure that all the data is not stored to the same place. Instead data should exist in multiple vendors' servers. This achieves market mobility and protects against server outages. Instead of having an entire copy of a file in multiple places, more economical way is to use erasure coding to store a certain part of a file to each vendor. In this method one file is stored in *n* parts over *n* vendors so that *k* pieces of *n* are needed for file reconstruction (k<n). While this method admittedly increases storage costs, the benefits of distributed cloud storage make it a valid solution to vendor lock-in problem. Distributed cloud storage has a number of different implementations that tackle different problems and should be chosen based on customers' needs.

Distributed cloud storage has been identified as a solution to the lock-in problem but it feels wrong that third party has to fix a huge flaw in the field. It feels like a quick bandage fix over the inherent problems of cloud storage services, lack of portability and interoperability. This paper proposes that future research be dedicated to developing and enforcing standards for CSPs to make data migration easier.

Vendor lock-in is a major deterrent to the adoption of cloud storage. It reduces portability and reliability which are considered one of the benefits of cloud storage. It makes cloud storage more expensive and risky and makes it seem like a less attractive option for data storage. In the future, industry needs to enforce standards for interoperability between CSPs. While this will increase competition in pricing, it will grow the industry as a whole by speeding up the adoption of cloud storage services.

REFERENCES

Abu-Libdeh, H., Princehouse, L., & Weatherspoon, H. (2010). RACS: a case for cloud storage diversity. *In Proceedings of the 1st ACM symposium on Cloud computing,* p. 229-240

Amazon S3 (2015). retrieved on 22.6.2015 from: http://aws.amazon.com/s3/details/

Armbrust, M. Fox, A. Griffith, R. Joseph, A.D., Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., Stoica, I. & Zaharia, M. (2009). Above the Clouds: A Berkeley View of Cloud Computing. *Electrical Engineering and Computer Sciences University of California at Berkeley*

Bowers, K. D., Juels, A., & Oprea, A. (2009, November). HAIL: a high-availability and integrity layer for cloud storage. *In Proceedings of the 16th ACM conference on Computer and communications security* p. 187-198

Brodkin, J. (2008). Loss of customer data spurs closure of online storage service 'The Linkup'. *Network World.*

Couts, A. (2012). "Upload at your own risk: most cloud storage services offer no data guarantee". Digital Trends.

Chua, C. E. H., Lim, W. K., Sia, S. K., Soh, C. (2008). "Threat-Balancing in Vendor Transition". International Research Workshop on IT Project Management 2008. Paper 3.

Crump, G. (2014). Retrieved on 12.6.2015 from: http://searchcloudstorage.techtarget.com/tip/Switching-cloud-storage-service-providers-comes-at-a-cost

EMC Corporation (2014). MANAGING STORAGE: TRENDS, CHALLENGES, AND OPTIONS (2013-2014). Retrieved on 12.6.2015 from: https://education.emc.com/content/_common/docs/articles/Managing_Storage_Trends_Challenges_and_Options_2013_2014.pdf

GoGrid (2015). Retrieve on 22.6.2015 from: http://www.gogrid.com/products/storage/cloud

Google Cloud Storage (2015). Retrieved on 22.6.2015 from: https://cloud.google.com/storage/

Govindarajan, A., Lakshmanan (28.5.2010). Overview of Cloud Standards. *Computer Communications and Networks 2010*, 77-89

Janssen, C. (2015). Defintion – What does Cloud Storage mean?. Retrieved on 23.6.2015 from: http://www.techopedia.com/definition/26535/cloud-storage.

Leavitt, N. (2009). Is cloud computing really ready for prime time?. *Computer, (1),* p.15-20.

Mell, P., Grance, T. (2011). The NIST definition of cloud computing.

Moyle, E. (2012). Cloud computing vendor lock-in: Avoiding security pitfalls. Techtarget. Retrieved on 12.6.2015 from: http://searchcloudsecurity.techtarget.com/tip/Cloud-computing-vendor-lock-in-Avoiding-security-pitfalls

Nasuni (2014). Bulk Data Migration in the Cloud [White paper]. Retrieved on 1.7.2015, from Nasuni press release: http://www.nasuni.com/resource/white-paper-bulk-data-migration-in-the-cloud/

Okoli, C. & Schabram, K. (2010). A Guide to Conducting a Systematic Literature Review of Information Systems Research. *Available at SSRN 1954824.*

Opara-Martins, J., Sahandi, R., Tian, F. (2014). Critical Review of Vendor Lock-in and Its Impact on Adoption of Cloud Computing. *Information Society (i-Society), 2014 International Conference on* , vol., no., pp.92,97, 10-12

Petcu, D. (2011). Portability and Interoperability between Clouds: Challenges and Case Study. *Lecture Notes in Computer Science Volume 6994*, p. 62-74

Pearson, S., Benameur, A. (2010). Privacy, Security and Trust Issues Arising from Cloud Computing. *Cloud Computing Technology and Science (CloudCom), 2010 IEEE Second International Conference on vol., no.,* p.693-702

Rabbetts, A. (30.12.2013). Cloud supplier lock-in – our experience. *ComputerWeekly*

Razavian, S.M., Khani, H. Nasser, Y., & Ghassemi, F. (2013). An analysis of vendor lock-in problem in cloud storage. *Computer and Knowledge Engineering (ICCKE), 2013 3th International eConference*

Satzger, B., Hummer, W., Inzinger, C., Leitner, P., & Dustdar, S. (2013) Winds of Change: From Vendor Lock-In to the Meta Cloud. *Internet Computing, IEEE , vol.17, no.1,* p.69-73

Schnjakin, M., & Meinel, C. (2013). Implementation of cloud-raid: A secure and reliable storage above the clouds. In *Grid and Pervasive Computing* (pp. 91-102). Springer Berlin Heidelberg.

Slamanig, D., & Hanser, C. (2012). On cloud storage and the cloud of clouds approach. *Internet Technology And Secured Transactions, 2012 International Conference for , vol., no.,* p.649,655

Taneja, A. (April 2012). Dealing with cloud storage service providers: Avoiding vendor lock-in. Techtarget. Retrieved on 7.7.2015 from: http://searchcloudstorage.techtarget.com/tip/Dealing-with-cloud-storage-service-providers-Avoiding-vendor-lock-in

The Linux Information Project. (2006). Website http://www.linfo.org/vendor_lockin.html

Toosi, A. N., Calheiros, R. N., Buuya, R. (July 2014). Interconnected Cloud Computing Environments: Challenges, Taxonomy, and Survey. *ACM Computing Surveys (CSUR), Volume 47 Issue 1.*

Wang, C., Ren, K., Lou, W., & Li, J. (2010). Toward publicly auditable secure cloud data storage services. *Network, IEEE, 24(4),* 19-24

Whitten, D. (2010). Adaptability in IT Sourcing: The Impact of Switching Costs. *Global Sourcing of Information Technology and Business Processes,* pp. 202-216

Wu, J., Ping, L., Ge, X., Wang, Y., Fu, J. (June 2010). "Cloud Storage as the Infrastructure of Cloud Computing," *Intelligent Computing and Cognitive Informatics (ICICCI), 2010 International Conference on* , vol., no., pp. 380,383, 22-23.