Controlling Data in the Cloud: Outsourcing Computation without Outsourcing Control
It is to be declared that all the work done in this document is solely and explicitly by the author of this document.

The list of literatures, whose text has been adapted, is mentioned in the document with reference.

This term paper is based on a scientific paper with the title *Controlling Data in the Cloud: Outsourcing Computation without Outsourcing Control*, written by: Richard Chow, Philippe Golle, Markus Jakobsson, Elaine Shi, Jessica Staddon, Ryusuke Masuoka, Jesus Molina.

June 8th 2012
Wirayudha Rohandi
# Table of Contents

Table of Contents ........................................................................................................... i
Table of Figures ................................................................................................................ ii

1. Introduction ......................................................................................................................... 1
   1.1. Cloud Computing ............................................................................................................. 1
   1.2. Cloud Computing Control ............................................................................................... 3

2. Fear of The Cloud ............................................................................................................... 4
   2.1. Traditional Security .......................................................................................................... 4
       2.1.1. VM-level Attacks ........................................................................................................ 4
       2.1.2. Cloud Provider Vulnerabilities ................................................................................. 5
       2.1.3. Phishing Cloud Provider ........................................................................................... 5
       2.1.4. Expanded Network Attack Surface ......................................................................... 5
       2.1.5. Authentication and Authorization .............................................................................. 6
       2.1.6. Forensics in the Cloud ............................................................................................... 6
   2.2. Availability ....................................................................................................................... 6
       2.2.1. Uptime ....................................................................................................................... 6
       2.2.2. Single Point of Failure ............................................................................................... 6
       2.2.3. Assurance of Computational Integrity ........................................................................ 7
   2.3. Third-party Data Control ................................................................................................. 7
       2.3.1. Due Diligence ............................................................................................................. 7
       2.3.2. Auditability ................................................................................................................ 7
       2.3.3. Contractual Obligations ............................................................................................. 7
       2.3.4. Cloud Provider Espionage ......................................................................................... 7
       2.3.5. Data Lock-in ............................................................................................................. 7
       2.3.6. Transitive Nature ...................................................................................................... 8

3. New Problems ..................................................................................................................... 9
   3.1. Data Location & Privacy ................................................................................................. 9
   3.2. Cost-effective Defense of Availability .......................................................................... 10
   3.3. Increased Authentication Demands ............................................................................... 10
   3.4. Mash-up Authorization ................................................................................................. 10

4. New Directions .................................................................................................................... 11
   4.1. Information-centric Security ......................................................................................... 11
   4.2. High-Assurance Remote Server Attestation ................................................................. 11
   4.3. Privacy-Enhanced Business Intelligence ...................................................................... 11

5. Conclusion .......................................................................................................................... 13

6. References .......................................................................................................................... 14
# Table of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1</td>
<td>Cloud Computing Logical Diagram</td>
<td>2</td>
</tr>
<tr>
<td>Figure 2</td>
<td>Control of IT Infrastructure</td>
<td>3</td>
</tr>
<tr>
<td>Figure 3</td>
<td>General Challenges/Issues of Cloud</td>
<td>4</td>
</tr>
<tr>
<td>Figure 4</td>
<td>Multi-tenant Architectures</td>
<td>5</td>
</tr>
<tr>
<td>Figure 5</td>
<td>Cloud Type Based on Location</td>
<td>5</td>
</tr>
<tr>
<td>Figure 6</td>
<td>Single Point of Failure</td>
<td>6</td>
</tr>
</tbody>
</table>
1. Introduction

Over the last couple of years the word “Cloud Computing” became really famous. Cloud technology became more mature and more affordable than before, with 3 service models, cloud computing made huge impact on IT services. SaaS or Software as a Service on the top tier, PaaS or Platform as a Service on second tier and IaaS or Infrastructure as a Service are the 3 models, which can serve businesses really well.

The two factor which makes cloud technology could change traditional concept of computing are cost and scalability. This two factors help a lot of startup businesses or small to medium businesses to run their daily business without investing so much on IT infrastructure, in this case company’s data center.

With affordable cost and relatively easy to scale computing power and storage comes a serious problem, which is the control of the data. In this term paper I would like to address this particular topic, how can we outsource the computational power and storage without compromising control over the data.

Lack of control in the cloud is the major worry. One aspect of control is transparency in the cloud implementation, somewhat contrary to the original promise of cloud computing in which the cloud implementation is not relevant. Transparency is needed for regulatory reasons and to ease concern over the potential for data breaches. Because of today’s perceived lack of control, larger companies are testing the waters with smaller projects and less sensitive data. In short, the potential of the cloud is not being realized.

1.1. Cloud Computing

What is cloud computing? Everyone in the technology world is talking about it and a lot of people in the business world are asking the same question, “What is cloud computing, and what does it mean for my business?”

Most IT departments are forced to spend a significant portion of their time on frustrating implementation, maintenance, and upgrade projects that too often don’t add significant value to the company’s bottom line. Increasingly, IT teams are turning to cloud computing technology to minimize the time spent on lower-value activities and allow IT to focus on strategic activities with greater impact on the business.

The fundamental cloud computing infrastructure has won over the CIOs of some of the world’s largest organizations—these once-skeptical executives never looked back after experiencing first-hand the host of benefits delivered by cloud computing technology.

---

1 Salesforce, http://www.salesforce.com/cloudcomputing/
In this document, we explore the benefits and considerations of cloud computing, focusing on its potential for controlling data in the cloud while outsourcing computation without outsourcing control. Cloud computing offers a range of advantages, including:

- **Proven Web-services integration.** By their very nature, cloud-computing technology is much easier and quicker to integrate with your other enterprise applications (both traditional software and cloud computing infrastructure-based), whether third-party or homegrown.

- **World-class service delivery.** Cloud computing infrastructures offer much greater scalability, complete disaster recovery, and impressive uptime numbers.

- **No hardware or software to install:** a 100% cloud computing infrastructure. The beauty of cloud computing technology is its simplicity... and in the fact that it requires significantly fewer capital expenditures to get up and running.

- **Faster and lower-risk deployment.** You can get up and running in a fraction of the time with a cloud computing infrastructure. No more waiting months or years and spending millions of dollars before anyone gets to log into your new solution. Your cloud computing technology applications are live in a matter of weeks or months, even with extensive customization or integration.

- **Support for deep customizations.** Some IT professionals mistakenly think that cloud computing technology is difficult or impossible to customize extensively, and therefore is not a good choice for complex enterprises. The cloud computing infrastructure not only allows deep customization and

---


Controlling Data in the Cloud: Outsourcing Computation without Outsourcing Control
application configuration, it preserves all those customizations even during upgrades. And even better, cloud computing technology is ideal for application development to support your organization’s evolving needs.

- **Empowered business users.** Cloud computing technology allows on-the-fly, point-and-click customization and report generation for business users, so IT doesn’t spend half its time making minor changes and running reports.

- **Automatic upgrades that don’t impact IT resources.** Cloud computing infrastructures put an end to a huge IT dilemma: If we upgrade to the latest-and-greatest version of the application, we’ll be forced to spend time and resources (that we don’t have) to rebuild our customizations and integrations. Cloud computing technology doesn’t force you to decide between upgrading and preserving all your hard work, because those customizations and integrations are automatically preserved during an upgrade.

### 1.2. Cloud Computing Control

Before company decides to move to cloud environment, it is a good practice to know who is in control of the data, or the possibility of the provider to provide assurances. This practice makes company know what are the stakes if they move to cloud, how much do they want to put sensitive data on the cloud, which role do they want to have, and a lot of other factors.

![Control of IT Infrastructure](http://srmsblog.burtongroup.com/2009/06/cloud-computing-who-is-in-control.html)

**Figure 2 - Control of IT Infrastructure**

---

2. Fear of The Cloud

Based on IDC survey, companies are afraid the most of security issues on the cloud environment, with significant effect of application performance on the second position.

![General Challenges/Issues of Cloud](image)

This section will discuss about 15 domains, which are preventing companies from taking advantage of the cloud. These 15 domains will be divided by 3 main categories:

- Traditional security
- Availability
- Third-party data control

2.1. Traditional Security

These concerns involve computer and network intrusions or attacks that will be made possible or at least easier by moving to the cloud. Cloud providers respond to these concerns by arguing that their security measures and processes are more mature and tested than those of the average company.

2.1.1. VM-level Attacks

Potential vulnerabilities in the hypervisor or VM technology used by cloud vendors are a potential problem in multi-tenant architectures.

---

The problem with multi-tenant architectures is, if the host’s operating system has a hole, a guest’s operating system could compromise the data of the other tenants.

2.1.2. Cloud Provider Vulnerabilities
These could be platform level, such as an SQL-injection or cross-site scripting vulnerability.

2.1.3. Phishing Cloud Provider
Phishing is attempting to acquire information (and sometimes, indirectly, money) such as usernames, passwords, and credit card details by masquerading as a trustworthy entity in an electronic communication, in this case phishing comes from the cloud provider.

2.1.4. Expanded Network Attack Surface
The cloud user must protect the infrastructure used to connect and interact with the cloud, a task complicated by the cloud being outside the firewall in many cases.

---

6 http://en.wikipedia.org/wiki/Phishing
7 http://blog.qualitance.ro/2011/05

Controlling Data in the Cloud: Outsourcing Computation without Outsourcing Control
2.1.5. Authentication and Authorization

The enterprise authentication and authorization framework does not naturally extend into the cloud. How does a company meld its existing framework to include cloud resources? Furthermore, how does an enterprise merge cloud security data (if even available) with its own security metrics and policies?

2.1.6. Forensics in the Cloud

“Traditional digital forensic methodologies permit investigators to seize equipment and perform detailed analysis on the media and data recovered. The likelihood therefore, of the data being removed, overwritten, deleted or destroyed by the perpetrator in this case is low. More closely linked to a CC environment would be businesses that own and maintain their own multi-server type infrastructure, though this would be on a far smaller scale in comparison. However, the scale of the cloud and the rate at which data is overwritten is of concern.”

2.2. Availability

These concerns center on critical applications and data being available.

2.2.1. Uptime

As with the Traditional Security concerns, cloud providers argue that their server uptime compares well with the availability of the cloud user’s own data centers.

2.2.2. Single Point of Failure

Cloud services are thought of as providing more availability, but perhaps not – there are more single points of failure and attack.

---

8 CLOIDIFIN. http://community.zdnet.co.uk/blog/0,1000000567,2000625196b,00.htm?new_comment.
2.2.3. Assurance of Computational Integrity
Can an enterprise be assured that a cloud provider is faithfully running a hosted application and giving valid results? For example, Stanford’s Folding@Home project gives the same task to multiple clients to reach a consensus on the correct result.

2.3. Third-party Data Control
The legal implications of data and applications being held by a third party are complex and not well understood. There is also a potential lack of control and transparency when a third party holds the data. Part of the hype of cloud computing is that the cloud can be implementation independent, but in reality regulatory compliance requires transparency into the cloud.

2.3.1. Due Diligence
If served a subpoena or other legal action, can a cloud user compel the cloud provider to respond in the required time-frame? A related question is the provability of deletion, relevant to an enterprise’s retention policy: How can a cloud user be guaranteed that data has been deleted by the cloud provider?

2.3.2. Auditability
Audit difficulty is another side effect of the lack of control in the cloud. Is there sufficient transparency in the operations of the cloud provider for auditing purposes? Currently, this transparency is provided by documentation and manual audits. One popular auditing guideline is the SAS 70, which defines guidelines for auditors to assess internal controls, for instance controls over the processing of sensitive information. SOX and HIPAA are other well-known regulations. US government agencies generally need to follow guidelines from FISMA, NIST, and FIPS.

2.3.3. Contractual Obligations
One problem with using another company's infrastructure besides the uncertain alignment of interests is that there might be surprising legal implications.

2.3.4. Cloud Provider Espionage
This is the worry of theft of company proprietary information by the cloud provider. For example, Google Gmail and Google Apps are examples of services supported by a private cloud infrastructure. Corporate users of these services are concerned about confidentiality and availability of their data.

2.3.5. Data Lock-in
How does a cloud user avoid lock-in to a particular cloud-computing vendor? The data might itself be locked in a proprietary format, and there are also issues with training and processes.
2.3.6. Transitive Nature

Another possible concern is that the contracted cloud provider might itself use subcontractors, over whom the cloud user has even less control, and who also must be trusted. One example is the online storage service called The Linkup, which in turn used an online storage company called Nirvanix. The Linkup shutdown after losing sizeable amounts of customer data, which some say was the fault of Nirvanix.
3. New Problems

In this section the journal outline new problem areas in security that arise from cloud computing. These problems may only become apparent after the maturation and more widespread adoption of cloud computing as a technology.

3.1. Data Location & Privacy\textsuperscript{10}

Data in a cloud computing environment has to exist on physical servers somewhere in the world and the physical location of those servers is important under many nations’ laws.

This is especially important for companies that do business across national boundaries, as different privacy and data management laws apply in different countries.

For example, the European Union places strict limits on what data can be stored on its citizens and for how long. Many banking regulators also require customers’ financial data to stay in their home country. Many compliance regulations require that data not be intermixed with other data, such as on shared servers or databases.

In another example, Canada is concerned about its public sector projects being hosted on U.S.-based servers because under the U.S. Patriot Act, it could be accessed by the U.S. government.

Some of the larger cloud providers (e.g. Microsoft, Google) have recognised this issue and are starting to allow customers to specify the location of their data.

Another data issue to address is: What happens to your data in a legal entanglement?

What if you miss paying a bill, or decide not to pay a bill for various reasons, like dissatisfaction with the service? Do you lose access to your data?

This is something that can be addressed at the contract stage to ensure that the right safeguards are be put in place to prevent a provider from withholding access to your data.

3.1. Cheap Data and Data Analysis

The rise of cloud computing has created enormous data sets that can be monetized by applications such as advertising. Google, for instance, leverages its cloud infrastructure to collect and analyze consumer data for its advertising network. Collection and analysis of data is now possible cheaply, even for companies lacking Google’s resources. What is the impact on privacy of abundant data and cheap data-

mining? Because of the cloud, attackers potentially have massive, centralized databases available for analysis and also the raw computing power to mine these databases. For example, Google is essentially doing cheap data mining when it returns search results. How much more privacy did one have before one could be Googled?

3.2. Cost-effective Defense of Availability

Availability also needs to be considered in the context of an adversary whose goals are simply to sabotage activities. The damages are not only related to the losses of productivity, but extend to losses due to the degraded trust in the infrastructure, and potentially costly backup measures. The cloud computing model encourages single points of failure. It is therefore important to develop methods for sustained availability (in the context of attack), and for recovery from attack. The latter could operate on the basis of minimization of losses, required service levels, or similar measures.

3.3. Increased Authentication Demands

The development of cloud computing may, in the extreme, allow the use of thin clients on the client side. Rather than a license purchased and software installation on the client side, users will authenticate in order to be able to use a cloud application. There are some advantages in such a model, such as making software piracy more difficult and giving the ability to centralize monitoring. It also may help prevent the spread of sensitive data on untrustworthy clients.

Thin clients result in a number of opportunities related to security, including the paradigm in which typical users do not have to worry about the risks of any actions – their security is managed by the cloud, which maintains the software they run. This architecture stimulates mobility of users, but increases the need to address authentication in a secure manner. In addition, the movement towards increased hosting of data and applications in the cloud and lesser reliance on specific user machines is likely to increase the threat of phishing and other abusive technologies aimed at stealing access credentials, or otherwise derive them, e.g., by brute force methods.

3.4. Mash-up Authorization

As adoption of cloud computing grows, we are likely to see more and more services performing mash-ups of data. This development has potential security implications, both in terms of data leaks, and in terms of the number of sources of data a user may have to pull data from – this, in turn, places requirements on how access is authorized for reasons of usability. While centralized access control may solve many of these problems, that may not be possible – or even desirable.
4. New Directions

This section is for discussing the solution of the problems, which are raised on previous section.

4.1. Information-centric Security

In order for enterprises to extend control to data in the cloud, the proposal is shifting from protecting data from the outside (system and applications which use the data) to protecting data from within.

This self-protection requires intelligence be put in the data itself. Data needs to be self-describing and defending, regardless of its environment. Data needs to be encrypted and packaged with a usage policy. When accessed, data should consult its policy and attempt to re-create a secure environment using virtualization and reveal itself only if the environment is verified as trustworthy (using Trusted Computing). Information-centric security is a natural extension of the trend toward finer, stronger, and more usable data protection.

4.2. High-Assurance Remote Server Attestation

Lack of transparency is discouraging businesses from moving their data to the cloud. Data owners wish to audit how their data is being handled at the cloud, and in particular, ensure that their data is not being abused or leaked, or at least have an unalterable audit trail when it does happen. Currently customers must be satisfied with cloud providers using manual auditing procedures like SAS-70.

A promising approach to address this problem is based on Trusted Computing. Imagine a trusted monitor installed at the cloud server that can monitor or audit the operations of the cloud server. The trusted monitor can provide “proofs of compliance” to the data owner, stating that certain access policies have not been violated. To ensure integrity of the monitor, Trusted Computing also allows secure bootstrapping of this monitor to run beside (and securely isolated from) the operating system and applications.

The monitor can enforce access control policies and perform monitoring/auditing tasks. To produce a “proof of compliance”, the code of the monitor is signed, as well as a “statement of compliance” produced by the monitor. When the data owner receives this proof of compliance, it can verify that the correct monitor code is run, and that the cloud server has complied with access control policies.

4.3. Privacy-Enhanced Business Intelligence

A different approach to retaining control of data is to require the encryption of all cloud data. The problem is that encryption limits data use. In particular searching and
indexing the data becomes problematic. For example, if data is stored in clear-text, one can efficiently search for a document by specifying a keyword. This is impossible to do with traditional, randomized encryption schemes. State-of-the-art cryptography may offer new tools to solve these problems.

While in many cases more research is needed to make these cryptographic tools sufficiently practical for the cloud, we believe they present the best opportunity for a clear differentiator for cloud computing since these protocols can enable cloud users to benefit from one another’s data in a controlled manner. In particular, even encrypted data can enable anomaly detection that is valuable from a business intelligence standpoint. For example, a cloud payroll service might provide, with the agreement of participants, aggregate data about payroll execution time that allows users to identify inefficiencies in their own processes. Taking the vision even further, if the cloud service provider is empowered with some ability to search the encrypted data, the proliferation of cloud data can potentially enable better insider threat detection (e.g. by detecting user activities outside of the norm) and better data loss prevention (DLP) (e.g. through detecting anomalous content).

Apart from ensuring privacy, applied cryptography may also offer tools to address other security problems related to cloud computing. For example, in proofs of retrievability the storage server can show a compact proof that it is correctly storing all of the client’s data.
5. Conclusion

Cloud computing is not only a trend these days, but also cloud computing is really helping companies to do their daily business more efficient. But before a company decides to move to a cloud system, it should look at all the consequences and all the models available for the cloud environment.

The solutions proposed by the journal, I think, are still too early, we cannot implement them at the moment, because cloud technology is still new and we need to wait a little bit more, we need to let the technology approach the majority, because by then everyone will use this standard or approach. If we implement these proposals today, the overhead would be too big. If everyone were using thing approaches, the solutions technology would be cheap and affordable.

Regarding the cloud type solutions, if the company’s data are not required to be confidential, a company does not have to use private cloud environment, it can use public cloud environments.

Most of the startup businesses do not have big budget especially for cloud infrastructure, so if they are on tight budget, they can use combination of public cloud and private cloud environment or traditional in-house server, of course they will suffer on performance and hard to expand, but again to put on public space, the risk exposing the company’s data is still high. On the other hand, big companies will have enough money to create their private cloud environment, since their best interest are the security and confidential of their data.
6. References


CLOIDIFIN, http://community.zdnet.co.uk/blog/0,1000000567,2000625196b,00.htm?new_comment.


Controlling Data in the Cloud: Outsourcing Computation without Outsourcing Control