Towards Analyzing Data Security Risks in Cloud Computing Environments

by

Amit Sangroya, Saurabh Kumar, Jaideep Dhok, Vasudeva Varma

in

International Conference on Information Systems, Technology, and Management (ICISTM 2010)

Bangkok, Thailand,

Report No: IIIT/TR/2010/6

Centre for Software Engineering Research Lab
International Institute of Information Technology
Hyderabad - 500 032, INDIA
March 2010

Towards Analyzing Data Security Risks in Cloud Computing Environments

Amit Sangroya, Saurabh Kumar, Jaideep Dhok, and Vasudeva Varma

International Institute of Information Technology,
Hyderabad, India
{amit_s,saurabh,jaideep}@research.iiit.ac.in, vv@iiit.ac.in

Abstract. There is a growing trend of using cloud environments for ever growing storage and data processing needs. However, adopting a cloud computing paradigm may have positive as well as negative effects on the data security of service consumers. This paper primarily aims to highlight the major security issues existing in current cloud computing environments. We carry out a survey to investigate the security mechanisms that are enforced by major cloud service providers. We also propose a risk analysis approach that can be used by a prospective cloud service for analyzing the data security risks before putting his confidential data into a cloud computing environment.

1 Introduction

In a cloud computing environment, the underlying computing infrastructure is used only when it is needed. For example, in order to process a user request, a service provider can draw the necessary resources on-demand, perform a specific job and then relinquish the unneeded resources and often dispose them after the job is done. Contrary to traditional computing paradigms, in a cloud computing environment, data and the application is controlled by the service provider [1,2,3,4,5]. This leads to a natural concern about data safety and also its protection from internal as well as external threats. Despite of this, advantages such as *On demand infrastructure, pay as you go, reduced cost of maintenance, elastic scaling etc.* are compelling reasons for enterprises to decide on cloud computing environments.

Usually, in a cloud computing paradigm, data storage and computation are performed in a single datacenter. There can be various security related advantages in using a cloud computing environment. However, a single point of failure can not be assumed for any data loss. As shown in Figure 1, the data may be located at several geographically distributed nodes in the cloud. There may be multiple points where a security breach can occur. Compared to a traditional in house computing, it might be difficult to track the security breach in a cloud computing environment.

In this paper, we present the advantages and disadvantages (in the context of data security) of using a cloud environment. We carry out a small survey on

S.K. Prasad et al. (Eds.): ICISTM 2010, CCIS 54, pp. 255–265, 2010.
© Springer-Verlag Berlin Heidelberg 2010

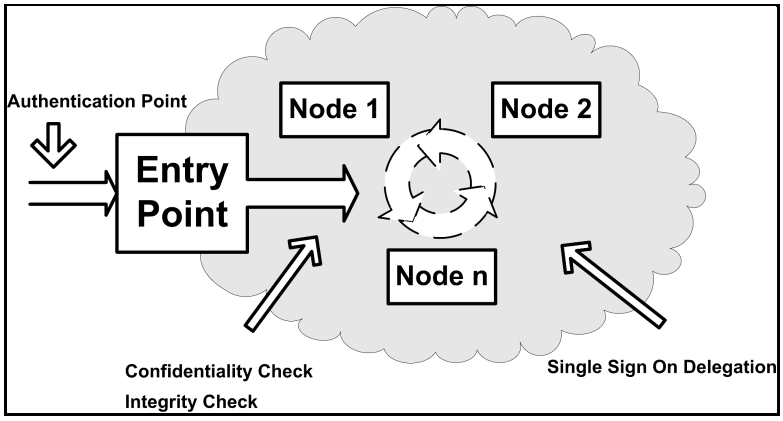

Fig. 1. Typical Data Security Checkpoints in a Cloud Computing Environment

major cloud service providers to investigate the prominent security issues. We investigate the security mechanisms that are used by major service providers. Our study supports that in the context of data security *trust* is a major element which is missing in the currently existing computing models. We believe a lack of trust management mechanism between the cloud service provider and users. Despite the fact that service providers use diverse mechanisms to retain high level of data security, we find a general lack of trust (in the context of confidential data) among the cloud service users.

In order to build a better trust mechanism, we present a risk analysis approach that can be primarily used by the perspective cloud users before putting their confidential data into a cloud. Our approach is based on the idea of trust model, principally used in distributed information systems [6,7]. We extend the general idea of trust management and present its use in analyzing the data security risks in cloud computing.

The contributions of this paper can be summarized as follows:

(a) We investigate the major security issues in cloud computing paradigms.
(b) We also carry out a survey of major cloud service providers to explore the security mechanisms in the context of security issues.
(c) Further, we also present a risk analysis approach that can be used by a prospective cloud service user to evaluate the risk of data security.

The structure of rest of the paper is as follows:

Section 2 provides insight into the security issues in the cloud computing environment from the perspective of service providers and service users. In section 3, we present our risk analysis approach from the perspective of a cloud user. In section 4, we present the related work. We present the limitations of our research and give some direction for future work in section 5.

2 Security Issues and Challenges

IaaS (Infrastructure as a Service), PaaS (Platform as a Service), and SaaS (Software as a Service) are three general models of cloud computing. Each of these models possess a different impact on application security [8]. However, in a typical scenario where an application is hosted in a cloud, two broad security questions that arises are:

- *How secure is the Data?*
- *How secure is the Code?*

Cloud computing environment is generally assumed as a potential cost saver as well as provider of higher service quality. *Security, Availability, and Reliability* are the major quality concerns of cloud service users. Gens et. al. [9], suggests that security in one of the prominent challenge among all other quality challenges.

2.1 Security Advantages in Cloud Environments

Current cloud service providers operate very large systems. They have sophisticated processes and expert personnel for maintaining their systems, which small enterprizes may not have access to. As a result, there are many direct and indirect security advantages for the cloud users. Here we present some of the key security advantages of a cloud computing environment:

Data Centralization: In a cloud environment, the service provider takes care of storage issues and small business need not spend a lot of money on physical storage devices. Also, cloud based storage provides a way to centralize the data faster and potentially cheaper. This is particularly useful for small businesses, which cannot spend additional money on security professionals to monitor the data.

Incident Response: IaaS providers can put up a dedicated forensic server that can be used *on demand* basis. Whenever a security violation takes place, the server can be brought online. In some investigation cases, a backup of the environment can be easily made and put onto the cloud without affecting the normal course of business.

Forensic Image Verification Time: Some cloud storage implementations expose a cryptographic check sum or hash. For example, Amazon S3 generates MD5 (Message-Digest algorithm 5) hash automatically when you store an object [10]. Therefore in theory, the need to generate time consuming MD5 checksums using external tools is eliminated.

Logging: In a traditional computing paradigm by and large, logging is often an afterthought. In general, insufficient disk space is allocated that makes logging either non-existent or minimal. However, in a cloud, storage need for standard logs is automatically solved.

2.2 Security Disadvantages in Cloud Environments

In spite of security advantages, cloud computing paradigm also introduces some key security challenges. Here we discuss some of these key security challenges:

Data Location: In general, cloud users are not aware of the exact location of the datacenter and also they do not have any control over the physical access mechanisms to that data. Most well-known cloud service providers have datacenters around the globe. Some service providers also take advantage of their global datacenters. However, in some cases applications and data might be stored in countries, which can judiciary concerns. For example, if the user data is stored in X country then service providers will be subjected to the security requirements and legal obligations of X country. This may also happen that a user does not have the information of these issues.

Investigation: Investigating an illegitimate activity may be impossible in cloud environments. Cloud services are especially hard to investigate, because data for multiple customers may be co-located and may also be spread across multiple datacenters. Users have little knowledge about the network topology of the underlying environment. Service provider may also impose restrictions on the network security of the service users.

Data Segregation: Data in the cloud is typically in a shared environment together with data from other customers. Encryption cannot be assumed as the single solution for data segregation problems. In some situations, customers may not want to encrypt data because there may be a case when encryption accident can destroy the data.

Long-term Viability: Service providers must ensure the data safety in changing business situations such as mergers and acquisitions. Customers must ensure data availability in these situations. Service provider must also make sure data security in negative business conditions like prolonged outage etc.

Compromised Servers: In a cloud computing environment, users do not even have a choice of using physical acquisition toolkit. In a situation, where a server is compromised; they need to shut their servers down until they get a previous backup of the data. This will further cause availability concerns.

Regulatory Compliance: Traditional service providers are subjected to external audits and security certifications. If a cloud service provider does not adhere to these security audits, then it leads to a obvious decrease in customer trust.

Recovery: Cloud service providers must ensure the data security in natural and man-made disasters. Generally, data is replicated across multiple sites. However, in the case of any such unwanted event, provider must do a complete and quick restoration.

Security Issues in Virtualization

Full Virtualization and Para Virtualization [11,12] are two kinds of virtualization in a cloud computing paradigm. In full virtualization, entire hardware architecture is replicated virtually. However, in para virtualization, an operating system

is modified so that it can be run concurrently with other operating systems. VMM (Virtual Machine Monitor), is a software layer that abstracts the physical resources used by the multiple virtual machines. The VMM provides a virtual processor and other virtualized versions of system devices such as I/O devices, storage, memory, etc.

VMM Instance Isolation ensures that different instances running on the same physical machine are isolated from each other. However, current VMMs do not offer perfect isolation. Many bugs have been found in all popular VMMs that allow escaping from VM (Virtual machine). Vulnerabilities have been found in all virtualization softwares, which can be exploited by malicious users to bypass certain security restrictions or/and gain escalated privileges. Below are few examples for this:

(a) Vulnerability in Microsoft Virtual PC and Microsoft Virtual Server could allow a guest operating system user to run code on the host or another guest operating system [13].
(b) Vulnerability was found in VMware's shared folders mechanism that grants users of a guest system read and write access to any portion of the host's file system including the system folder and other security-sensitive files.
(c) Vulnerability in Xen can be exploited by "root" users of a guest domain to execute arbitrary commands [12].

2.3 Survey

We carry out a small survey of major cloud service providers to investigate the security mechanisms to overcome the security issues discussed in this paper. We consider ten major cloud service providers. These providers provide their services in all major areas of cloud computing, including SaaS, PaaS and IaaS. Table 1 shows the list of service providers that we studied in this survey. In order to analyze the complete state of art of security in cloud computing, the survey needs to be more exhaustive. However, due to the fact that the scope of our work is not just to explore the state of art but to look at the major factors that affect security in cloud computing. Therefore we have intentionally not considered other cloud service providers in this survey.

In table 2, we present the results of the survey that depicts the current state of security mechanisms. Information given in table 2 is based on the information available online at the official websites of these providers.

Table 1. Major Cloud Service Providers

Service Provider Type	Names
IaaS	Amazon EC2, Amazon S3, GoGrid
PaaS	Google App Engine, Microsoft Azure Services, Amazon Elastic Map Reduce
SaaS	Salesforce, Google Docs

Table 2. Summary of Security Mechanisms by Major Cloud Service Providers

Security Issue	Results
Password Recovery	90% are using standard methods like other common services, while 10% are using sophisticated techniques.
Encryption Mechanism	40% are using standard SSL encryption, while 20% are using encryption mechanism but at an extra cost. 40% are using advance methods like HTTPS access also.
Data Location	70% have their datacenters located in more than one country, while 10% are located at a single location. 20% are not open about this issue.
Availability History	In 40% there is a reported downtime alongwith a result in data loss, while in 60% cases data availability is good.
Proprietary/Open	Only 10% providers have open mechanism.
Monitoring Services	70% are providing extra monitoring services, while 10% are using automatic techniques. 20 % are not open about this issue.

3 Risk Analysis Approach

The cloud computing service providers use various security mechanisms to ensure that all the security risks are fully taken care of. However, there are two broad questions:

- *How to estimate the risk to data security before putting a job into the cloud?* and
- *How to ensure customers that their data and programs are safe in provider's premises?*

If a cloud service user is able to estimate the risk of his data security then he can have a level of trust with the service provider. If there is a high risk about the data security then it leads to a decrease in trust and vice-versa.

3.1 Need of a Risk Analysis Approach

The service users need a clear communication about the methods adopted by the service providers to maintain security. Current security technology provides us with some capability to build a certain level of trust in cloud computing. For example, SSL (Secure Socket Layer), digital signatures, and authentication protocols for proving authentication and access control methods for managing authorization. However, these methods cannot manage the more general concept of *Trustworthiness*. SSL, for instance, cannot on its own prove that if a communication between server and multiple hosts is secure. Also, as we discussed in section 1, there are multiple points of failures in a cloud environment.

Current security technology is lacking the complementary tool for managing trust effectively. Gambetta et. al. [14] define trust as *"trust (or, symmetrically,*

distrust) is a particular level of the subjective probability with which an agent will perform a particular action, both before [we] can monitor such action (or independently of his capacity of ever to be able to monitor it) and in a context in which it affects [our] own action". Based on this definition, we can say that trust is a subjective property and is affected by those actions that we cannot monitor.

Three kinds of trust models have been discussed in distributed computing [6]:

- *Direct Trust;*
- *Transitive Trust; and*
- *Assumptive Trust*

In a cloud computing paradigm, where the data and programs cross the organizational boundaries, transitive trust and assumptive trust can be crucial for certain type of applications. A direct trust model in a cloud exists in cloud computing environments, when there is a common trust entity that performs all original entity authentications and the generation of credentials that are bound to specific entities.

A key difference with other models is that the direct trust model does not allow the delegation of original entity authentication. And every relying party must use this entity directly for all validation processes. An example of such type of trust entity is the use of PKI[1] based authentication where a root certification authority (CA) do all kinds of trust relationships. The responsibility of secure data transfer lies in the hands of the certificating authority.

3.2 Risk Assessment Using Trust Matrix

Although no single unit of measure is adequate to the definition of trust, several dependent variables (such as data cost), can be used to describe it. Based upon the prominent security factors discussed in section 2, we build a trust matrix to analyze the data risk. To build the trust matrix, a number of heuristics can be used for selecting the security parameters. However, a simple way to select the security factors is to prioritize them based on subjective opinion and select two most important parameters. We select following two trust variables to build the trust matrix:

(a) *Data Cost*
(b) *Provider's History*

The reason for selecting these trust variables is explained here:

In cloud environment, data can be assigned a cost by the users based on the criticality of the data. The data criticality needs to be computed by the service users. There may be multiple factors that affect the data criticality. For example,

[1] The Public Key Infrastructure (PKI) is a set of hardware, software, people, policies, and procedures needed to create, manage, distribute, use, store, and revoke digital certificates. PKI is an arrangement that binds public keys with respective user identities by means of a certificate authority (CA) [15,16].

confidential business data can be critical and therefore we can assign it a higher cost as compared to less critical data.

Similarly, service provider history can be a possible parameter to estimate the risk. History includes a provider's profile of past services. If users are dissatisfied with a particular service, they can record their experience. If a service provider do not possess a good history of data security (e.g. there is a past record of security failures), then it may also decrease the trust factor. However, other variables can also be used for building the trust matrix. Some of these variables can be Encryption support, Service Cost, Monitoring support etc.

Variable Parameters
Along with trust variables, few parameters used in measuring trust can be applied to fine-tune these trust variables. The parameters which we choose in this category are:

(a) *Data Location*
(b) *Regulatory Compliance*

As we have explained in section 2, data located at the sites which are geographically or politically sensitive would likely to have lower trust than other locations. Similarly, if a service provider is assuring the customers using a centralized regulating authority, it will lead to an increase in trust level among the service users.

We make use of variable parameters as a support mechanism in the trust matrix. It is used as a validation factor that provides a support in the risk analysis.

3.3 Risk Analysis

We capture the relationship using a trust matrix where the axes represents the variables used. The variables used should be meaningfully related to each other. Figure 2, represents an example trust matrix with area representing the Low Risk/High Trust zone and, High Risk/Low Trust zone. This can be explained as:

- x axis represents the data cost.
- y axis represents the service provider's history. and
- z axis represents the data location.

Now, it is obvious that a high data cost with poor service provider history combining with a very sensitive location will result in a higher risk/lower trust.

High trust zone signifies the region of high trust. It can specify the security risk for the current transactions and also for future transactions with that service provider. Similarly, *low trust zone* signifies the region of low trust.

As a risk preventive approach, we also define here a *trust action*, which can be taken as part of a preventive or reactive measure. For example, an added level of authentication and/or verification can be used for the activities which are related to the low trust zone. We have used these variables in a common

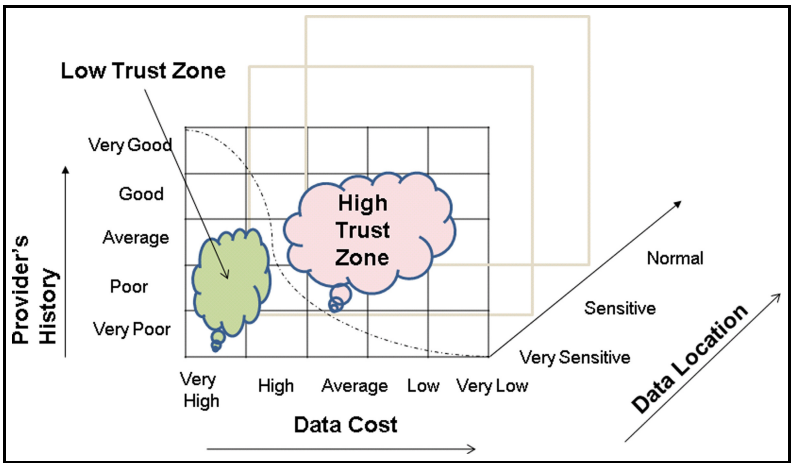

Fig. 2. A Trust Matrix for Risk Analysis

cloud computing scenario, where we have some past statistics about the service provider. The method has been used to measure the trust and will be used for all future transactions. Based on this method, we were able to define the trust actions, for all future transactions with the service provider.

4 Related Work

A survey conducted by IDC (International Data Corporation) suggests that cloud services are still in the early adoption phase. There is a long list of issues cloud service providers need to address. The survey has rated security as the most prominent concern [9]. Buyya [17], provide a survey on current state of the art in cloud computing and identify key challenges that must be addressed in order to make cloud computing a reality.

Cachin et. al. [18], in their survey, give insight into the well known cryptographic tools for providing integrity and consistency for data stored in clouds. The security solutions explored and discussed by them are keeping a local copy of the data, use of hash tree, protocols such as Proofs of Retrievability (POR), and Proofs of Data Possessions (PDP), Digital Signatures etc. These solutions still require a testing on some live data to validate their suitability and ease of use. A whitepaper by AWS (Amazon Web Services) discusses physical security, backups, and certifications in their context [10]. Similarly, other providers such as Google, Microsoft etc. have discussed the security issues in cloud computing [19,20].

Heiser and Nicolett [21] have identified seven prominent risks that customers must assess in order to utilize cloud computing infrastructure. In addition to these seven issues, we have also identified several other major issues that must be addressed by the cloud service providers. These issues include data storage,

server security, privileged user access, and data portability. We also present virtualization specific security issues in detail.

Manchala et. al. [7] have build a trust model in a distributed computing paradigm. To the best of our knowledge, none of the work so far, give a direction to address the security challenges, specifically in cloud environments. Despite the fact that, there are solutions to address the prominent security issues, a mechanism to measure the security risk from the perspective of a service user is strongly needed. Trust models have been studied in distributed information systems [6,7]. Adopting some of the ideas of trust modeling, our work identifies a key set of trust variables and a resulting trust matrix, based on security issues in cloud computing.

5 Conclusion

In an emerging discipline, like cloud computing, security needs to be analyzed more frequently. With advancement in cloud technologies and increasing number of cloud users, data security dimensions will continuously increase. In this paper, we have analyzed the data security risks and vulnerabilities which are present in current cloud computing environments.

The most obvious finding to emerge from this study is that, there is a need of better trust management. We have built a risk analysis approach based on the prominent security issues. The security analysis and risk analysis approach will help service providers to ensure their customers about the data security. Similarly, the approach can also be used by cloud service users to perform risk analysis before putting their critical data in a security sensitive cloud.

At present, there is a lack of structured analysis approaches that can be used for risk analysis in cloud computing environments. The approach suggested in this paper is a first step towards analyzing data security risks. This approach is easily adaptable for automation of risk analysis.

References

1. Hayes, B.: Cloud Computing. Communications ACM 51, 9–11 (2008)
2. Amazon elastic compute cloud (2008), http://aws.amazon.com/ec2/
3. Twenty Experts Define Cloud Computing (2008), http://cloudcomputing.syscon.com/read/612375_p.htm
4. Llanos, D.R.: Review of Grid Computing Security by Anirban Chakrabarti. Queue 5, 45 (2007)
5. Weiss, A.: Computing in the Clouds. NetWorker 11, 16–25 (2007)
6. Andert, D., Wakefield, R., Weise, J.: Trust Modeling for Security Architecture Development (2002), http://www.sun.com/blueprints
7. Manchala, D.W.: E-Commerce Trust Matrix and Models (2000)
8. John, H.: Security Guidance for Critical Areas of Focus in Cloud Computing (2009), http://www.cloudsecurityalliance.org/guidance/ (Accessed 2 July 2009)
9. Gens, F.: IT Cloud Services User Survey, part 2: Top Benefits and Challenges (2008)

10. Overview of Security Processes (2008)
11. Rose, R.: Survey of System Virtualization Techniques (2004), http://www.robertwrose.com/vita/rose-virtualization.pdf
12. Xen Multiple Vulnerabilities (2007), http://secunia.com/advisories/26986/
13. Microsoft Security Bulletin MS07-049 (2007), http://www.microsoft.com/technet/security/bulletin/ms07-049.mspx
14. Diego, G.: Can we trust Trust? Oxford:Trust Making and Breaking Cooperative Relations (1990)
15. Two Factor Authentication, http://en.wikipedia.org/wiki/
16. Public Key, http://en.wikipedia.org/wiki/Public_key_certificate
17. Buyya, R., Yeo, C.S., Venugopal, S., Broberg, J., Brandic, I.: Cloud Computing and Emerging IT Platforms: Vision, Hype, and Reality for delivering Computing as the 5th Utility. Future Generation Computer Systems 25, 599–616 (2009)
18. Cachin, C., Keider, I., Shraer, A.: Trusting The Cloud. IBM Research, Zurich Research laboratory (2009)
19. Google App Engine (2008), http://appengine.google.com
20. Microsoft Live Mesh (2008), http://www.mesh.com
21. Brodkin, J.: Seven Cloud Computing Security Risks (2008), http://www.gartner.com/DisplayDocument?id=685308