ITM Web of Conferences **17**, 03028 (2018) https://doi.org/10.1051/itmconf/20181703028
WCSN 2017

Data security and risk assessment in cloud computing

Jing Li [*] *and Qinyuan Li* [b]

State Grid Zhejiang Electric Power Research Institute, Hangzhou, China
[b]+86 13819177903, liqinyuan@zj.sgcc.com.cn

Abstract. Cloud computing has attracted more and more attention as it reduces the cost of IT infrastructure of organizations. In our country, business Cloud services, such as Alibaba Cloud, Huawei Cloud, QingCloud, UCloud and so on are gaining more and more uses, especially small or median organizations. In the cloud service scenario, the program and data are migrating into cloud, resulting the lack of trust between customers and cloud service providers. However, the recent study on Cloud computing is mainly focused on the service side, while the data security and trust have not been sufficiently studied yet. This paper investigates into the data security issues from data life cycle which includes five steps when an organization uses Cloud computing. A data management framework is given out, including not only the data classification but also the risk management framework. Concretely, the data is divided into two varieties, business and personal information. And then, four classification levels (high, medium, low, normal) according to the different extent of the potential adverse effect is introduced. With the help of classification, the administrators can identify the application or data to implement corresponding security controls. At last, the administrators conduct the risk assessment to alleviate the risk of data security. The trust between customers and cloud service providers will be strengthen through this way.

1 Introduction

Cloud computing is growing more and more attraction in recent years, as it has transformed the way organizations approach IT, reducing the IT cost by the novel service models. Nowadays, many organizations are increasingly transferring their application and data on the cloud, though losing the control of the application and data [1]. By the transforming of control of the application and data, organizations can gain the benefits in both efficiency and effectiveness, as well as reduce the expenses in purchasing and sustaining the infrastructure, especially for the acceptance of the public cloud computing. Many reports show that the cloud computing market continues to grow explosively. According to reports from Gartner [2], in 2017, the worldwide public cloud services market grows 18 percent to

[*] Jing Li: 18606830186@163.com

overall $246.8 billion, in contrast with $209.2 billion in 2016. Projected to grow 36.8 percent in 2017 to reach $34.6 billion, IaaS (Infrastructure as a Service) will be the highest growth. However, the SaaS (Software as a Service) is anticipated to grow 20.1 percent to reach $46.3 billion (see Table 1.)

Table 1. Worldwide Public Cloud Services Forecast (Millions of Dollars)

	2016	***2017***	***2018***	***2019***	***2020***
Cloud Business Process Services (BPaaS)	40,812	43,772	47,556	51,652	56,176
Cloud Application Infrastructure Services (PaaS)	7,169	8,851	10,616	12,580	14,798
Cloud Application Services (SaaS)	38,567	46,331	55,143	64,870	75,734
Cloud Management and Security Services	7,150	8,768	10,427	12,159	14,004
Cloud System Infrastructure Services (IaaS)	25,290	34,603	45,559	57,897	71,552
Cloud Advertising	90,257	104,516	118,520	133,566	151,091
Total Market	211,261	248,858	289,839	334,743	385,375

[a.] Source: Gartner (February 2017)

As SaaS products become more sophisticated, including HCM and CRM, the SaaS market will show a slight slowdown in development during the years to come. Despite this, SaaS will still be the second largest section of the global cloud services market.

But for security experts, the cloud has a big dilemma. How do you use the benefits of the cloud while maintaining security control over your organization's applications and data? Determining whether increased risk is really worth agility and economic benefit is a matter of balance. Maintaining data management is the most important for the success of the cloud. Ten years ago, enterprise data typically resides on its own server within the organization's physical infrastructure, enterprise data center, and was able to separate confidential data into individual physical servers. Nowadays, in the cloud and virtualization, data may be under logical control of the institution, but physically it resides in the infrastructure belonged to and maintained by another organization.

NIST SP 800-145[3] defines four deployment models: private, community, public and hybrid cloud. Private cloud and public cloud are the most important and popular deployment models. In private cloud, the infrastructure is provided for dedicated use by a single organization offering service to multiple consumers. However, in public cloud, the infrastructure is designed to provide open use for the public. Different with private cloud, the infrastructure in public cloud is owned, managed, and operated by other cloud service provider, or a third party, but not the organization itself [4]. From the perspective of the network perimeter, public cloud is available and open to the public—is accessible via the internet. On the other hand, the private cloud is only open limitedly, behind the firewall which is under control by the organization. Besides, the private cloud is operated by organization itself. How to secure the application and data without direct control in public cloud is paramount for large-scale application.

Due to the lack of trust between users of the cloud service and cloud service providers, it is important that customers evaluate the security of the data. Risk evaluation is a standard practice that allows organizations to identify, demonstrate and follow other strategies to

avoid data loss in the cloud. In this article, we first analyze the data security issues that cloud computing uses from the life cycle of the data. Data security issues are inherent as the lack of trust between cloud service users and cloud service provider. And then, we will give the data classification for reference on the data security. At last, a risk assessment framework will be described.

2 Data security in life cycle

Security content in the cloud resembles traditional security issue and is embodied in all stages of the life cycle. However, as a result of cloud virtualization and multi-tenancy, the content of data security in the cloud has its own unique features.

This section analyzes data security problems in the data life cycle [5] [6]. The data lifecycle represents the whole procedure from data production to dumping. The data life cycle is divided into five segments, as shown in Figure 1.

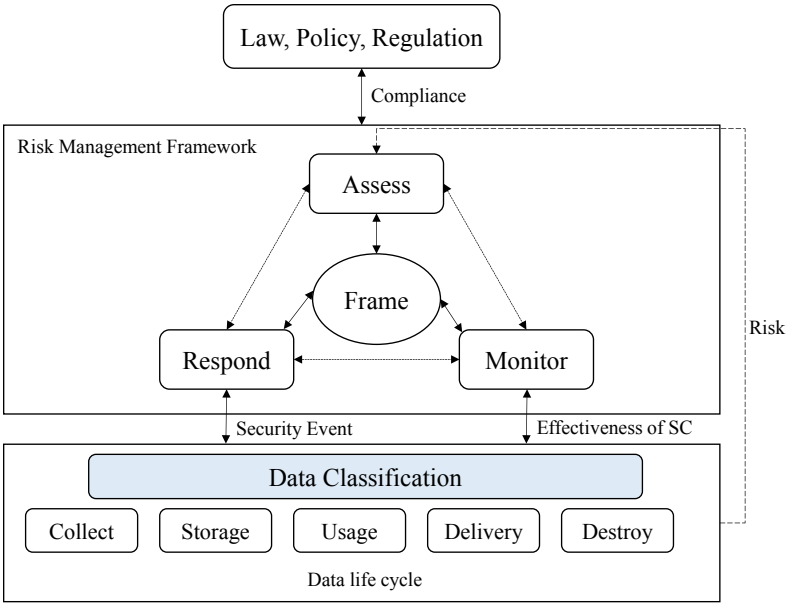

Figure 1. Data Management Framework

2.1 Collect

There are three security concerns in data collection phase: ownership, categorization and quality.

Ownership: In the traditional IT models, the data is stored in the local memory, organizations have the ownership. However, But in cloud models, the data ownership becomes a problem as the fact that the data is migrated Especially, for the personally identifiable information(PII), In particular, the data owner has the right to know what information of personally identifiable information is collected.

Categorization: Categorization will help organizations allocate their resources, prioritize the selection and placement of security controls, and ensure that sensitive systems meet baseline security standards.

Quality: Data may be collected from different sources with different format and quality. From a security perspective, securing the data should start from the collection phase and ensure the raw data is trusted and accurate.

2.2 Storage

Three aspects of information security must be considered in the data that stored in the cloud: confidentiality, integrity, and accessibility (CIA). Data encryption is a usual way to ensure data security. Besides, encryption algorithms and key management should be considered for the validity of encryption.

Data integrity should also be taken into account besides data confidentiality. If users place some GB (or more) data on cloud storage, how can they control the integrity of the data? As a fast and flexible particularity of cloud computing resources, users do not know the location where data is stored. Outbound or cloud storage requires users to use bandwidth and time. Some cloud venders, for example, Amazon, a fee is required for the data transfer. A method to directly certify the consistency of data in cloud storage with no uploading data after downloading data is a big issue. In the traditional IT models, the external attacks become the major threat to availability of data. But for the cloud models, other aspects that threat availability of data arise: (1) The availability of cloud computing services; (2) Cloud providers will always be stably operating? (3) Will the backup be provided in the cloud storage services?

2.3 Usage

Data encryption can be performed on static data using uncomplicated data storage services, such as Amazon S3. But for static data used in cloud applications such as PaaS or SaaS models, it is often unsuitable to encrypt data. Static data used in cloud applications is often not encrypted because data encryption can cause problems with index creation and queries. Almost all processed data is not encrypted in the cloud as well as in traditional IT environments. Because of the multi-party nature of cloud computing models, information processed by cloud applications and data from other users are stored jointly. A serious threat to data security can be posed by unencrypted data in this process. The utilization of personal data is more difficult. Owners of private data must know and be sure that the use of private information is consistent with the objective for which it was collected and if the personal data can be disclosed to third parties such as cloud service providers.

2.4 Delivery

The dissemination of data expands the use of data and complicates data rights. A data owner can transmit data in general, and other organizations can access data with no ratification of the data owner. So once data is delivered, the data owner must consider whether the user retains the original protection and usage restrictions. As for the distribution of private data, besides data authorization, the fineness of data distribution and transformation is also important. The granularity of the distribution is determined by the content strategy and the granularity of the segmentation. Data conversion refers to the separation of confidential information from the original data, which makes the data independent of the owner of the data.

2.5 Destroy

Because of the physical features of storage medium, the deleted data still exists and can be regained, which leads to unintentional disclosure of confidential information. Therefore, if the storage medium is not used any more, the information stored on the medium must be completely deleted.

3 Data classification

Data classification is a very efficient method to protect the data according to its importance and sensitivity. Document NIST Special Publication 800-60 (Guide for Mapping Types of Information and Information Systems to Security Categories) provides a guide for organizations to conduct data classification early in the design of information systems. With the help of the classification, the administrators can identify the application or data to implement corresponding security controls.

In this section, we introduce four classification levels (high, medium, low, normal) according to the different extent of the potential side influence on organizational operations, institutional assets, or individuals.

- Confidentiality: A scarcity of confidentiality is the **unauthorized leak** of data.
- Integrity: A scarcity of integrity is the **unauthorized alteration or damage** of data
- Availability: A scarcity of availability is the **commotion of access** to or use of data.

Table 2. Data Classification

No.	**Classification**	
	Business Information	***Personal Information***
High Sensitivity	Mid-long term development programs and special plans etc.	Identification card number, passport number etc.
Medium Sensitivity	Important indicator of marketing operations etc.	Email address or contents etc.
Low Sensitity	Company address book etc.	Payment recodes, etc.
Normal Sensitity	Public information (could be found from library or internet)	

3.1 High sensitivity

If the missing of CIA of data is anticipated to have a severe or destructive side effect, the classification of data is high sensitivity. Serious or destructive results indicate, for instance: (i) The severe decrease or loss of mission abilities results in the organization losing its ability to fulfill its primary function. (ii) Serious damage to plan assets. (iii) Significant financial loss. or (iv) r Serious or catastrophic damage to individuals.

Examples of high sensitivity data includes but not limited:

1) Core business information: Core business information is the distinctly important information for organization, the loss of which is very important on the mission, financial, reputation and so on. For example:

a) The mid-long term development programs and special plans

b) Safety protection scheme

c) key information for important infrastructure

d) core network topology

2) Personal sensitive information: Personal sensitive information is closely related to personal rights or interests, the disclosure or abuse of which may endanger property safety

and life, adverse impact on personal reputation, leading to discriminatory treatment and other sensitive information. For example:

a) Identification card number, passport number, etc.

b) Credit/debit card information and bank account information

c) Protected Health Information

d) Electronic signatures, biometric information, password information, and private encryption keys

e) Criminal background

3) Industry-specific sensitive information: From perspective of national security, some Industry-specific sensitive information such as oil, gas, coal and power plant related information, should be protected strickly. For example:

a) Oil and gas production information

b) Petrochemical industry important production materials import plan and foreign exchange amount

c) Distribution of Urban Power Network Pipeline

d) Power transmission and transformation equipment reliability paremeter

3.2 Medium sensitity

When the loss of data of CIA is anticipated to have a serious side effect, the classification of data is medium sensitivity. A severe side effect indicates that, for instance : (i) The significant decrease of mission abilities to fulfill its primary function and the effectiveness of the functions is reduced (ii) substantial damage to institutional assets; (iii) substantial financial loss; or (iv) huge damage to individuals that involves nether mortality nor fatal injuries.

Examples of medium sensitivity data includes but not limited:

1) Important business information: Important business information is the information the loss of which will result in serious adverse impact on the mission, financial, reputation and so on. For example:

a) important indicator of marketing operations

b) operation monitoring core indicators and operational data

c) computer software source code

d) Important business agreements or contracts and the relevant information

2) Personal normal information: Personal information b(Personal information is information that is recorded electronically or otherwise, which can be used alone or in combination with other information to identify citizens, including but not limited to citizens' names, birth dates, identity document numbers, personal biometric information, telephone number and so on.) is divided into personal sensitive information and personal normal information. Thus, if personal information is not sensitive, it may be normal information. Generally, personal normal information cannot identify a person alone, but in combination with other information togather.

a) Email address or contents

b) phone number, home address

c) Information used to validate identity, such as name, date of birth, mother’s name, etc.

d) other information not labeled as sensitive information (network user accout, race, ethnicity, marital status).

3.3 Low sensitity

When the loss of data of CIA is anticipated to have a limited side effect, the classification of data is low sensitivity. A limited side effect indicates that, for instance:(i)a decline in the

capability of carrying out tasks to an extent and interval that the organization has the ability to fulfill its foremost functions, as well as noticeably reduce the effectiveness of the functions (ii) negligible damage to institutional assets; (iii) trivial financial loss; or (iv) minor damage to individuals.

1) Normal business information: Normal business information is the information the loss of which will result in limited adverse impact on the mission, financial, reputation and so on. For example:

a) Company address book

b) final accounts and audit report computer of projects

c) Important business agreements or contracts and the relevant information

d) employee performance review information

2) Personal service information: Personal service information is the data and content information that the organization collects in the service process with personal privacy attributes. For example:

a) payment recodes, such as gas, electricity, mobile phone bill, etc.

b) device information, such as Meter, electricity collection terminal

c) service content inforamtion, such as business charges, user price program, user price strategy, etc.

3.4Normal sensitity

If data is expected to be public, the classification of data is normal sensitivity. Most of the information from the internet is normal sensitive information.

4 Risk assessment framework

Risk management activity is carried out to address risk throughout the organization as a comprehensive and organization-wide activity. Risk evaluation is one of the pivotal components in the organization-wide risk management process (RMP), which is defined in NIST Special Publication 800-39, Managing Information Security Risk: Organization, Mission, and Information System View [8]. In [8], RMP includes four components: (a) framework risk; (b) evaluating risk; (c) Responding to risk; and (d) Monitoring risk. Fig.1 demonstrates the risk evaluation framework among the four components.

Risk framework is the principal step to evaluate risk, accentuating how organizations mount risks or build risk settings. And the “settings” portrays the environment of cloud or any other information system. It is truly onerous to set up a practical and high-efficiency framework, since the organizations not only make appropriate evaluations but also identify the risk limits.

Table 3. Components in RMF

Components	***Introduction/Purpose***
Framework	(1)To establish a risk context. (2)To engender a risk management scheme
Evaluating	(1) To identify the threats and vulnerabilities; (2) To identify the harm
Responding	To develop alternative countermeasures
Monitoring	To certify that risk response measures are put in to practice effectively

Risk evaluation is the secondary step after risk framing, which addresses the risk evaluation problems. It consists of two aspects: one is identifying the imperilments and vulnerabilities; the other is identifying the damage. The imperilments include both internal and external factors. And the damage means the adverse event when the adversaries manipulate some vulnerability successfully. In [9], risk is a function illustrating the probability of a imperilment event's incidence and potential side influence should the event take place. Generally speaking, risk evaluation process consists of four steps: (a) preparing for the evaluation; (b) conducting the evaluation; (c) communicating evaluation results; and (d) maintaining the evaluation.

Built upon the result of risk evaluation, risk responding components will respond to the risk to solve the problems and mediate the side effect as soon as possible. Organizations will carry out risk responses according to strategies and controls properly.

Nevertheless, risk cannot be eradicated absolutely and may change over time. The two ways that organizations monitor risk over time and evaluate the risk on ongoing foundations are the two key components of risk monitoring. The objective of surveillance is to certify that risk response measures are put into use. Consistent surveillance can classify configuration-related variations to cloud and the environments of operation. After that whether the risk response measures are effectual of not can be determinate. Through this method, the risk could be upheld at a comparatively low level all through.

5 Conclusion

In this paper, we analyze the data security issues of utilizing cloud computing from data life cycle. As the characteristics of service hosting, data security issues are inherent as the lack of trust between cloud service users and cloud service provider. We give the data classification for reference from business Information and personal information. The rank of data is divided into four levels (high, medium, low, normal). At last, we give out a risk management framework to guide the data security management.

References

[1] V. Chang and M. Ramachandran, "Towards Achieving Data Security with the Cloud Computing Adoption Framework," in IEEE Transactions on Services Computing, vol. 9, no. 1, pp. 138-151, Jan.-Feb. 1 2016.
[2] Gartner Says Worldwide Public Cloud Services Market to Grow 18 Percent in 2017. https://www.gartner.com/newsroom/id/3616417
[3] SP 800-145, "The NIST Definition of Cloud Computing," NIST Special Publications, Sept. 2011.
[4] M. Henze, J. Hiller, O. Hohlfeld and K. Wehrle, "Moving Privacy-Sensitive Services from Public Clouds to Decentralized Private Clouds," 2016 IEEE International Conference on Cloud Engineering Workshop (IC2EW), Berlin, 2016, pp. 130-135.
[5] T. V. Sathyanarayana and L. M. I. Sheela, "Data security in cloud computing," 2013 International Conference on Green Computing, Communication and Conservation of Energy (ICGCE), Chennai, 2013, pp. 822-827.
[6] X. Yu and Q. Wen, "A View about Cloud Data Security from Data Life Cycle," 2010 International Conference on Computational Intelligence and Software Engineering, Wuhan, 2010, pp. 1-4.
[7] SP 800-60, "Volume I: Guide for Mapping Types of Information and Information Systems to Security Categories", NIST Special Publications, Aug. 2008.

[8] SP 800-39, "Managing information security risk organization, mission, and information system view," NIST Special Publications, Mar. 2010.
[9] SP 800-30 revision 1, "Guide for conducting risk assessments," NIST Special Publications, Sept. 2012.