Data Security and Privacy Protection Issues in Cloud Computing

Deyan Chen[1]
College of Information Science and Engineering
Northeastern University[1]
Shenyang, China
Email:chendeyan@neusoft.com

Hong Zhao[1,2]
Academy
Neusoft Corporation[2]
Shenyang, China

***Abstract*—It is well-known that cloud computing has many potential advantages and many enterprise applications and data are migrating to public or hybrid cloud. But regarding some business-critical applications, the organizations, especially large enterprises, still wouldn't move them to cloud. The market size the cloud computing shared is still far behind the one expected. From the consumers' perspective, cloud computing security concerns, especially data security and privacy protection issues, remain the primary inhibitor for adoption of cloud computing services. This paper provides a concise but all-round analysis on data security and privacy protection issues associated with cloud computing across all stages of data life cycle. Then this paper discusses some current solutions. Finally, this paper describes future research work about data security and privacy protection issues in cloud.**

Keywords-access control; cloud computing; cloud computing security; data segregation; data security; privacy protection.

I. INTRODUCTION

From initial concept building to current actual deployment, cloud computing is growing more and more mature. Nowadays many organizations, especially Small and Medium Business (SMB) enterprises, are increasingly realizing the benefits by putting their applications and data into the cloud. The adoption of cloud computing may lead to gains in efficiency and effectiveness in developing and deployment and save the cost in purchasing and maintaining the infrastructure.

Regarding definition of cloud computing model, the most widely used one is made by NIST as "*Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This cloud model promotes availability and is composed of five essential characteristics, three service models, and four deployment models.*"[1] The cloud computing model NIST defined has three service models and four deployment models. The three service models, also called SPI model, are: Cloud Software as a Service (SaaS), Cloud Platform as a Service (PaaS) and Cloud Infrastructure as a Service (IaaS). The four deployment models are: Private cloud, Community cloud, Public cloud and Hybrid cloud.

Compared with the traditional IT model, the cloud computing has many potential advantages. But from the consumers' perspective, cloud computing security concerns remain a major barrier for the adoption of cloud computing. According to a survey from IDCI in 2009, 74% IT managers and CIOs believed that the primary challenge that hinders them from using cloud computing services is cloud computing security issues [2]. Another survey carried out by Garter in 2009, more than 70% CTOs believed that the primary reason not to use cloud computing services is that there are data security and privacy concerns.

Although cloud computing service providers touted the security and reliability of their services, actual deployment of cloud computing services is not as safe and reliable as they claim. In 2009, the major cloud computing vendors successively appeared several accidents. Amazon's Simple Storage Service was interrupted twice in February and July 2009. This accident resulted in some network sites relying on a single type of storage service were forced to a standstill. In March 2009, security vulnerabilities in Google Docs even led to serious leakage of user private information. Google Gmail also appeared a global failure up to 4 hours. It was exposed that there was serious security vulnerability in VMware virtualization software for Mac version in May 2009. People with ulterior motives can take advantage of the vulnerability in the Windows virtual machine on the host Mac to execute malicious code. Microsoft's Azure cloud computing platform also took place a serious outage accident for about 22 hours. Serious security incidents even lead to collapse of cloud computing vendors. As administrators' misuse leading to loss of 45% user data, cloud storage vendor LinkUp had been forced to close.

Security control measures in cloud are similar to ones in traditional IT environment. As multi-tenant characteristic, service delivery models and deploy models of cloud computing,

Supported by Core Electronic Device, High General Chip and Basic Software program of China: 2011ZX01043-001-001
Supported by National Natural Science Foundation of China: 60803131
Supported by Electronic Information Industry Development Fund Project: "Multi-industries oriented Information Technology Services Knowledge Base System Development."
Supported by National basic research program of China (973): 2012CB724107

978-0-7695-4647-6/12 $26.00 © 2012 IEEE
DOI 10.1109/ICCSEE.2012.193

compared with the traditional IT environment, however, cloud computing may face different risks and challenges.

Traditional security issues are still present in cloud computing environments. But as enterprise boundaries have been extended to the cloud, traditional security mechanisms are no longer suitable for applications and data in cloud. Due to the openness and multi-tenant characteristic of the cloud, cloud computing is bringing tremendous impact on information security field:

(1) Due to dynamic scalability, service abstraction, and location transparency features of cloud computing models, all kinds of applications and data on the cloud platform have no fixed infrastructure and security boundaries. In the event of security breach, it's difficult to isolate a particular physical resource that has a threat or has been compromised.

(2) According to the service delivery models of cloud computing, resources cloud services based on may be owned by multiple providers. As there is a conflict of interest, it is difficult to deploy a unified security measures;

(3) As the openness of cloud and sharing virtualized resources by multi-tenant, user data may be accessed by other unauthorized users.

(4) As the cloud platform has to deal with massive information storage and to deliver a fast access, cloud security measures have to meet the need of massive information processing.

This paper describes data security and privacy protection issues in cloud. This paper is organized as follows: Section II gives a brief description of what exactly cloud computing security-related issues are. Section III discusses data security and privacy protection issues associated with cloud computing across all stages of data life cycle. Section IV shows current solutions for data security and privacy protection issues in cloud. Section V summarizes the contents of this paper. Section VI describes future research work.

II. CLOUD COMPUTING SECURITY ISSUES

A. Cloud Computing Security

Wikipedia [3] defines Cloud Computing Security as "*Cloud computing security (sometimes referred to simply as "cloud security") is an evolving sub-domain of computer security, network security, and, more broadly, information security. It refers to a broad set of policies, technologies, and controls deployed to protect data, applications, and the associated infrastructure of cloud computing.*" Note that cloud computing security referred to here is not cloud-based security software products such as cloud-based anti-virus, anti-spam, anti-DDoS, and so on.

B. Security Issues Associated with the Cloud

There are many security issues associated with cloud computing and they can be grouped into any number of dimensions.

According to Gartner [4], before making a choice of cloud vendors, users should ask the vendors for seven specific safety issues: Privileged user access, regulatory compliance, data location, data segregation, recovery, investigative support and long-term viability. In 2009, Forrester Research Inc. [5] evaluated security and privacy practices of some of the leading cloud providers (such as Salesforce.com, Amazon, Google, and Microsoft) in three major aspects: Security and privacy, compliance, and legal and contractual issues. Cloud Security Alliance (CSA) [6] is gathering solution providers, non-profits and individuals to enter into discussion about the current and future best practices for information assurance in the cloud. The CSA has identified thirteen domains of concerns on cloud computing security [7].

S. Subashini and V. Kavitha made an investigation of cloud computing security issues from the cloud computing service delivery models (SPI model) and give a detailed analysis and assessment method description for each security issue [8]. Mohamed Al Morsy, John Grundy and Ingo Müller explored the cloud computing security issues from different perspectives, including security issues associated with cloud computing architecture, service delivery models, cloud characteristics and cloud stakeholders [9]. Yanpei Chen, Vern Paxson and Randy H. Katz believed that two aspects are to some degree new and essential to cloud: the complexities of multi-party trust considerations, and the ensuing need for mutual auditability. They also point out some new opportunities in cloud computing security [10].

According to the SPI service delivery models, deployment models and essential characteristics of cloud, there are security issues in all aspects of the infrastructure including network level, host level and application level.

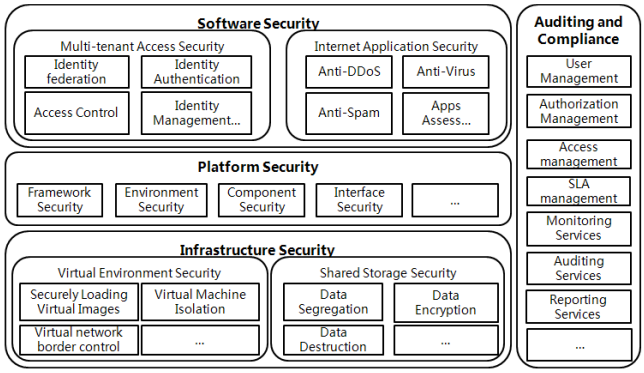

Figure 1. Cloud computing security architecture

III. DATA SECURITY AND PRIVACY PROTECTION ISSUES

The content of data security and privacy protection in cloud is similar to that of traditional data security and privacy protection. It is also involved in every stage of the data life cycle. But because of openness and multi-tenant characteristic of the cloud, the content of data security and privacy protection in cloud has its particularities.

The concept of privacy is very different in different countries, cultures or jurisdictions. The definition adopted by Organization for Economic Cooperation and Development (OECD) [11] is "*any information relating to an identified or*

identifiable individual (data subject)." Another popular definition provided by the American Institute of Certified Public Accountants (AICPA) and the Canadian Institute of Chartered Accountants (CICA) in the Generally Accepted Privacy Principles (GAPP) standard is "*The rights and obligations of individuals and organizations with respect to the collection, use, retention, and disclosure of personal information.*" Generally speaking, privacy is associated with the collection, use, disclosure, storage, and destruction of personal data (or personally identifiable information, PII). Identification of private information depends on the specific application scenario and the law, and is the primary task of privacy protection.

The next several sections analyze data security and privacy protection issues in cloud around the data life cycle.

A. Data Life Cycle

Data life cycle refers to the entire process from generation to destruction of the data. The data life cycle is divided into seven stages. See the figure below:

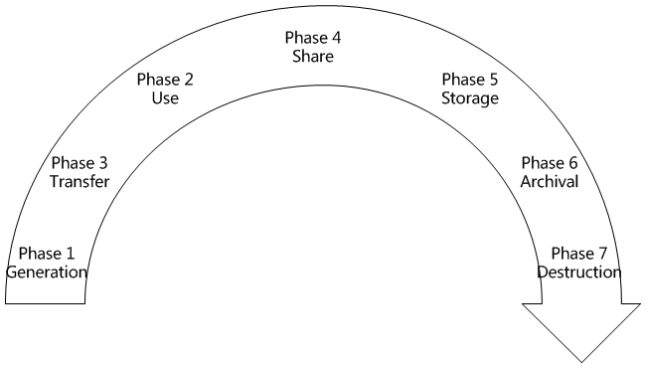

Figure 2. Data life cycle

B. Data Generation

Data generation is involved in the data ownership. In the traditional IT environment, usually users or organizations own and manage the data. But if data is to be migrated into cloud, it should be considered that how to maintain the data ownership.

For personal private information, data owners are entitled to know what personal information being collected, and in some cases, to stop the collection and use of personal information.

C. Transfer

Within the enterprise boundaries, data transmission usually does not require encryption, or just have a simple data encryption measure. For data transmission across enterprise boundaries, both data confidentiality and integrity should be ensured in order to prevent data from being tapped and tampered with by unauthorized users. In other words, only the data encryption is not enough. Data integrity is also needed to be ensured. Therefore it should ensure that transport protocols provide both confidentiality and integrity.

Confidentiality and integrity of data transmission need to ensure not only between enterprise storage and cloud storage but also between different cloud storage services. In other words, confidentiality and integrity of the entire transfer process of data should be ensured.

D. Use

For the static data using a simple storage service, such as Amazon S3, data encryption is feasible. However, for the static data used by cloud-based applications in PaaS or SaaS model, data encryption in many cases is not feasible. Because data encryption will lead to problems of indexing and query, the static data used by Cloud-based applications is generally not encrypted. Not only in cloud, but also in traditional IT environment, the data being treated is almost not encrypted for any program to deal with. Due to the multi-tenant feature of cloud computing models, the data being processed by cloud-based applications is stored together with the data of other users. Unencrypted data in the process is a serious threat to data security.

Regarding the use of private data, situations are more complicated. The owners of private data need to focus on and ensure whether the use of personal information is consistent with the purposes of information collection and whether personal information is being shared with third parties, for example, cloud service providers.

E. Share

Data sharing is expanding the use range of the data and renders data permissions more complex. The data owners can authorize the data access to one party, and in turn the party can further share the data to another party without the consent of the data owners. Therefore, during data sharing, especially when shared with a third party, the data owners need to consider whether the third party continues to maintain the original protection measures and usage restrictions.

Regarding sharing of private data, in addition to authorization of data, sharing granularity (all the data or partial data) and data transformation are also need to be concerned about. The sharing granularity depends on the sharing policy and the division granularity of content. The data transformation refers to isolating sensitive information from the original data. This operation makes the data is not relevant with the data owners.

F. Storage

The data in the cloud may be divided into: (1) The data in IaaS environment, such as Amazon's Simple Storage Service; (2) The data in PaaS or SaaS environment related to cloud-based applications.

The data stored in the cloud storages is similar with the ones stored in other places and needs to consider three aspects of information security: confidentiality, integrity and availability.

The common solution for data confidentiality is data encryption. In order to ensure the effective of encryption, there needs to consider the use of both encryption algorithm and key strength. As the cloud computing environment involving large amounts of data transmission, storage and handling, there also needs to consider processing speed and computational

efficiency of encrypting large amounts of data. In this case, for example, symmetric encryption algorithm is more suitable than asymmetric encryption algorithm.

Another key problem about data encryption is key management. Is who responsible for key management? Ideally, it's the data owners. But at present, because the users have not enough expertise to manage the keys, they usually entrust the key management to the cloud providers. As the cloud providers need to maintain keys for a large number of users, key management will become more complex and difficult.

In addition to data confidentiality, there also needs to be concerned about data integrity. When the users put several GB (or more) data into the cloud storage, they how to check the integrity of the data? As rapid elasticity feature of cloud computing resources, the users don't know where their data is being stored. To migrate out of or into the cloud storage will consume the user's network utilization (bandwidth) and an amount of time. And some cloud providers, such as Amazon, will require users to pay transfer fees. How to directly verify the integrity of data in cloud storage without having to first download the data and then upload the data is a great challenge. As the data is dynamic in cloud storage, the traditional technologies to ensure data integrity may not be effective.

In the traditional IT environment, the main threat of the data availability comes from external attacks. In the cloud, however, in addition to external attacks, there are several other areas that will threat the data availability: (1) The availability of cloud computing services; (2) Whether the cloud providers would continue to operate in the future? (3) Whether the cloud storage services provide backup?

G. *Archival*

Archiving for data focuses on the storage media, whether to provide off-site storage and storage duration. If the data is stored on portable media and then the media is out of control, the data are likely to take the risk of leakage. If the cloud service providers do not provide off-site archiving, the availability of the data will be threatened. Again, whether storage duration is consistent with archival requirements? Otherwise, this may result in the availability or privacy threats.

H. *Destruction*

When the data is no longer required, whether it has been completely destroyed? Due to the physical characteristics of storage medium, the data deleted may still exist and can be restored. This may result in inadvertently disclose of sensitive information.

IV. CURRENT SECURITY SOLUTIONS FOR DATA SECURITY AND PRIVACY PROTECTION

IBM developed a fully homomorphic encryption scheme in June 2009. This scheme allows data to be processed without being decrypted [12]. Roy I and Ramadan HE applied decentralized information flow control (DIFC) and differential privacy protection technology into data generation and calculation stages in cloud and put forth a privacy protection system called airavat [13]. This system can prevent privacy leakage without authorization in Map-Reduce computing process. A key problem for data encryption solutions is key management. On the one hand, the users have not enough expertise to manage their keys. On the other hand, the cloud service providers need to maintain a large number of user keys. The Organization for the Advancement of Structured Information Standards (OASIS) Key Management Interoperability Protocol (KMIP) is trying to solve such issues [14].

About data integrity verification, because of data communication, transfer fees and time cost, the users can not first download data to verify its correctness and then upload the data. And as the data is dynamic in cloud storage, traditional data integrity solutions are no longer suitable. NEC Labs's provable data integrity (PDI) solution can support public data integrity verification [15]. Cong Wang proposed a mathematical way to verify the integrity of the data dynamically stored in the cloud [16].

In the data storage and use stages, Mowbray proposed a client-based privacy management tool [17]. It provides a user-centric trust model to help users to control the storage and use of their sensitive information in the cloud. Munts-Mulero discussed the problems that existing privacy protection technologies (such as K anonymous, Graph Anonymization, and data pre-processing methods) faced when applied to large data and analyzed current solutions [18]. The challenge of data privacy is sharing data while protecting personal privacy information. Randike Gajanayake proposed a privacy protection framework based on information accountability (IA) components [19]. The IA agent can identify the users who are accessing information and the types of information they use. When inappropriate misuse is detected, the agent defines a set of methods to hold the users accountable for misuse.

About data destruction, U.S. Department of Defense (DoD) 5220.22-M (the National Industrial Security Program Operating Manual) shows two approved methods of data (destruction) security, but it does not provide any specific requirements for how these two methods are to be achieved [20]. The National Institute of Standards and Technology (NIST) Special Publication [21], 800-88, gives a "*Guidelines for Media Sanitization.*"

V. CONCLUSION

Although cloud computing has many advantages, there are still many actual problems that need to be solved. According to a Gartner survey about cloud computing revenues, market size for Public and Hybrid cloud is $59 billion and it will reach USD 149B by 2014 with a compound annual growth rate of 20[22]. The revenue estimation implies that cloud computing is a promising industry. But from another perspective, existing vulnerabilities in the cloud model will increase the threats from hackers.

According to service delivery models, deployment models and essential features of the cloud computing, data security and privacy protection issues are the primary problems that need to be solved as soon as possible. Data security and privacy issues

exist in all levels in SPI service delivery models and in all stages of data life cycle.

The challenges in privacy protection are sharing data while protecting personal information. The typical systems that require privacy protection are e-commerce systems that store credit cards and health care systems with health data. The ability to control what information to reveal and who can access that information over the Internet has become a growing concern. These concerns include whether personal information can be stored or read by third parties without consent, or whether third parties can track the web sites someone has visited. Another concern is whether web sites which are visited collect, store, and possibly share personal information about users. The key to privacy protection in the cloud environment is the strict separation of sensitive data from non-sensitive data followed by the encryption of sensitive elements.

According to the analysis for data security and privacy protection issues above, it is expected to have an integrated and comprehensive security solution to meet the needs of defense in depth. Regarding privacy protection, privacy data identification and isolation are the primary tasks. They should be considered during the design of cloud-based applications.

VI. FUTURE WORK

For data security and privacy protection issues, the fundamental challenges are separation of sensitive data and access control. Our objective is to design a set of unified identity management and privacy protection frameworks across applications or cloud computing services. As mobility of employees in organizations is relatively large, identity management system should achieve more automatic and fast user account provisioning and de-provisioning in order to ensure no un-authorized access to organizations' cloud resources by some employees who has left the organizations. Authorization and access control mechanisms should achieve a unified, reusable and scalable access control model and meet the need of fine-grained access authorization. Accountability based privacy protection mechanisms will achieve dynamical and real-time inform, authorization and auditing for the data owners when their private data being accessed.

ACKNOWLEDGMENT

We thank all sponsors in the footnote on the first page for funding this ongoing research project and all volunteers for their involving this research project. We would also like to thank the anonymous referees for their constructive and valuable comments.

REFERENCES

[1] Peter Mell, and Tim Grance, "The NIST Definition of Cloud Computing," Version 15, 10-7-09, http://www.wheresmyserver.co.nz/storage/media/faq-files/cloud-def-v15.pdf.

[2] Sun Cloud Architecture Introduction White Paper (in Chinese). http://developers.sun.com.cn/blog/functionalca/resource/sun_353cloudcomputing_chinese.pdf.

[3] Cloud computing security, http://en.wikipedia.org/wiki/Cloud_computing_security.

[4] Gartner: Seven cloud-computing security risks. InfoWorld. 2008-07-02. http://www.infoworld.com/d/security-central/gartner-seven-cloud-computing-security-risks-853.

[5] Cloud Security Front and Center. Forrester Research. 2009-11-18. http://blogs.forrester.com/srm/2009/11/cloud-security-front-and-center.html

[6] Cloud Security Alliance. http://www.cloudsecurityalliance.org.

[7] Cloud Security Alliance, Security Guidance for Critical Areas of Focus in Cloud Computing, V2.1, http://www.cloudsecurityalliance.org/guidance/csaguide.v2.1.pdf.

[8] S. Subashini, V.Kavitha. A survey on security issues in service delivery models of cloud computing. Journal of Network and Computer Applications 34(2011)1-11.

[9] Mohamed Al Morsy, John Grundy, Ingo Müller, "An Analysis of The Cloud Computing Security Problem," in Proceedings of APSEC 2010 Cloud Workshop, Sydney, Australia, 30th Nov 2010.

[10] Yanpei Chen, Vern Paxson, Randy H. Katz, "What's New About Cloud Computing Security?" Technical Report No. UCB/EECS-2010-5. http://www.eecs.berkeley.edu/Pubs/TechRpts/2010/EECS-2010-5.html

[11] "OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data," http://www.oecd.org/document/18/0,3343,en_2649_34255_1815186_1_1_1_1,00.html.

[12] "IBM Discovers Encryption Scheme That Could Improve Cloud Security, Spam Filtering," at http://www.eweek.com/c/a/Security/IBM-Uncovers-Encryption-Scheme-That-Could-Improve-Cloud-Security-Spam-Filtering-135413/.

[13] Roy I, Ramadan HE, Setty STV, Kilzer A, Shmatikov V, Witchel E. "Airavat: Security and privacy for MapReduce," In: Castro M, eds. Proc. of the 7th Usenix Symp. on Networked Systems Design and Implementation. San Jose: USENIX Association, 2010. 297.312.

[14] "OASIS Key Management Interoperability Protocol (KMIP) TC", http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=kmip.

[15] Zeng K, "Publicly verifiable remote data integrity," In: Chen LQ, Ryan MD, Wang GL, eds. LNCS 5308. Birmingham: Springer-Verlag, 2008. 419.434.

[16] Cong Wang, Qian Wang, Kui Ren, and Wenjing Lou, "Ensuring Data Storage Security in Cloud Computing," in Proceedings of the 17th International Workshop on Quality of Service.2009:1-9.

[17] Bowers KD, Juels A, Oprea A. Proofs of retrievability: Theory and implementation. In: Sion R, ed. Proc. of the 2009 ACM Workshop on Cloud Computing Security, CCSW 2009, Co-Located with the 16th ACM Computer and Communications Security Conf., CCS 2009. New York: Association for Computing Machinery, 2009. 43.54. [doi: 10.1145/1655008.1655015]

[18] Muntés-Mulero V, Nin J. Privacy and anonymization for very large datasets. In: Chen P, ed. Proc of the ACM 18th Int'l Conf. on Information and Knowledge Management, CIKM 2009. New York: Association for Computing Machinery, 2009. 2117.2118. [doi: 10.1145/1645953.1646333]

[19] Randike Gajanayake, Renato Iannella, and Tony Sahama, "Sharing with Care An Information Accountability Perspective," Internet Computing, IEEE, vol. 15, pp. 31-38, July-Aug. 2011.

[20] DoD, "National Industrial Security Program Operating Manual", 5220.22-M, February 28, 2006.

[21] Richard Kissel, Matthew Scholl, Steven Skolochenko, Xing Li, "Guidelines for Media Sanitization," NIST Special Publication 800-88, September 2006, http://csrc.nist.gov/publications/nistpubs/800-88/NISTSP800-88_rev1.pdf.

[22] Gartner DataQuest Forecast on Public Cloud Services DocID G00200833, June 2, 2010.