Data Security in Cloud

Mr. Pankaj Sareen

Assistant Professor, Department of Computer Applications
SPN College, Mukerian, Punjab
pankaj.sareen1480@gmail.com

Dr. Tripat Deep Singh

Assistant Professor, Department of Computer Applications
GNIMT Ludhiana

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. Cloud Computing promises lower costs, rapid scaling, easier maintenance, and service availability anywhere and anytime. But regarding some business-critical applications, the organizations, especially large enterprises, still wouldn't move them to Cloud. The market size the Cloud Computing share 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. A recent Microsoft survey found that "58 percent of the public and 86 percent of business leaders are excited about the possibilities of Cloud Computing. But more than 90 percent of them are worried about security, availability, and privacy of their data as it rests in the Cloud." A key challenge is how to ensure and build confidence that the Cloud can handle user data securely.

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. This paper would also help the readers to know about the various risks and threats to the data in the Cloud. A short discussion on DPaaS is also discussed in this paper. Data Centre Security and NIST Guidelines on Security and Privacy in Public Cloud Computing are also discussed in this paper. Finally, this paper describes future research work about data security and privacy protection issues in Cloud.

Keywords-Cloud Computing, Data Life Cycle, DPaaS, Security and Privacy Issues

I. INTRODUCTION

Cloud Computing is a model for enabling service user's ubiquitous, convenient and 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. Cloud Computing [1, 2] is promising access to computing facilities from any location, in an economical, adaptable and upgradable way.

In spite of the several advantages that Cloud Computing brings along with it, there are several concerns and issues which need to be solved before ubiquitous adoption of this computing paradigm happens:

First, in Cloud Computing, the user may not have the kind of control over his/her data or the performance of his/her applications that he/she may need, or the ability to audit or change the processes and policies under which he/she must work.

Second, the Cloud customers may risk losing data by having them locked into proprietary formats and may lose control over their data since the tools for monitoring who is using them or who can view them are not always provided to the customers. Data loss is, therefore, a potentially real risk in some specific deployments.

Third, the standards are immature and insufficient for handling the rapidly changing and evolving technologies of Cloud Computing. Therefore, one cannot just move applications to the Cloud and expect them to run efficiently.

Finally, there are latency and performance issues since the Internet connections and the network links may add to latency or may put constraint on the available bandwidth.

II. RESEARCH METHODOLOGY USED

A. *Objectives of the Study*

1) To provide a concise but all-round analysis on Data Security and Privacy Protection issues associated with Cloud Computing across all stages of data life cycle
2) To know about the various risks and threats to the data in the Cloud
3) To discuss about DPaaS(Data Protection as a service)
4) To discuss Data Centre Security and NIST Guidelines on Security and Privacy in Public Cloud Computing

B. *Research Design*

The research is Literature Based research. This paper involves a comprehensive study of the earlier work done in this area by reviewers. The major purpose of this research is to provide a concise but all-round analysis on Data Security and Privacy Protection issues associated with Cloud Computing across all stages of data life cycle

C. Data Collection

Secondary data is used for the study. Data will be collected from the secondary sources like National Institute of Standards and Technology (NIST) Cloud Computing, Cloud Security Alliance (CSA), and various Research Papers based upon the Data Security of Cloud Computing.

III. DATA SECURITY LIFE CYCLE

One of the biggest security concerns people have when moving to the Cloud is related to the problem of keeping data secure and confidential. In this respect, some particular problems arise: who can create data, where the data is stored, who can access and modify data, what happens when data is deleted, how the back-up is done, how the data transfer occurs, etc. All of this is known as data security lifecycle [3]

Data life cycle refers to the entire process from generation to destruction of the data. The data life cycle is divided into seven stages [4]. See the figure below:

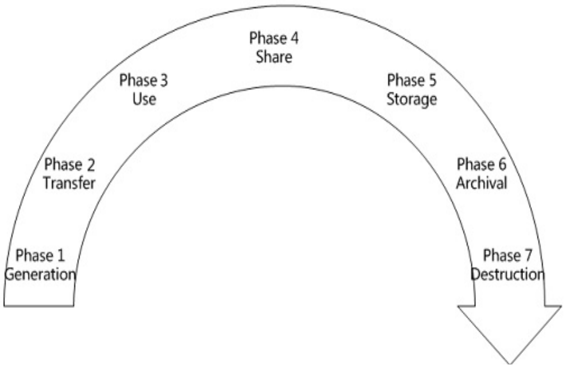

Fig.1 Data Security Life Cycle

A. 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.

B. Data 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.

C. Data Use

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.

D. Data Share

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.

E. Data Storage

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. As the Cloud Computing environment involves large amounts of data storage, there is need to consider processing speed and computational efficiency of encrypting large amounts of data. Key problem about data encryption is key management i.e. who is 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, the great challenge is how to directly verify the integrity of data in Cloud storage without having to first download the data and then upload the data?

The Concerns related to data availability are:

- The availability of Cloud Computing services;
- Whether the Cloud providers would continue to operate in the future?
- Whether the Cloud storage services provide backup?

F. Data 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.

G. Data 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. DATA SECURITY RISKS

The security risks [5] associated with each Cloud delivery model vary and are dependent on a wide range of factors including the sensitivity of information assets, Cloud architectures and security control involved in a particular Cloud environment. The various Data Security Risks in Cloud Computing are:

A. Privileged User Access

Once data is stored in the Cloud, the provider has access to that data and also controls access to that data by other entities (including other users of the Cloud and other third party suppliers). Maintaining confidentiality of data in the Cloud and limiting privileged user access can be achieved by at least one of two approaches by the data owner: first, encryption of the data prior to entry into the Cloud to separate the ability to store the data from the ability to make use of it; and second, legally enforcing the requirements of the Cloud provider through contractual obligations and assurance mechanisms to ensure that confidentiality of the data is maintained to required standards.

B. Data Location and Segregation

Data location and data segregation are of particular importance in the Cloud, given the disparate physical location of data and shared computing resources. Virtualization is one of a number of enabling technologies of Cloud Computing that it is a run-time method of segregation for processing data. Many of the security concerns and issues associated with virtualization are relevant in Cloud Computing. Security of data depends on having adequate security controls in each of the layers of the virtualized environment. In addition, secure deletion of memory and storage must be used to prevent data loss in a multi-tenant environment where systems are reused.

C. Data Disposal

Cloud services that offer data storage typically provide either guarantees or service-level objectives around high availability of that data. Cloud providers achieve this by keeping multiple copies of the data. Depending on the type of data hosted in the Cloud, customers may require providers to delete data in accordance with industry standards. Unless the Cloud architecture specifically limits the media on which data may be stored, customers may need to preclude their data from being transmitted in the Cloud.

D. Assessing the Security of a Third Party Cloud Provider

One of the most significant challenges for vendor Cloud customers in particular is assurance over the security controls of their Cloud provider. Customers are primarily concerned with the following issues:

1) *Defining security requirements:* The customers' information security requirements are derived from the organization's own policy, legal and regulatory obligations, and may carry through from other contracts or SLAs that the company has with its customers.

2) *Due diligence on Cloud service providers*: Prospective Cloud customers should undertake proper due-diligence on providers before entering into a formal relationship. Detailed due-diligence investigations can provide an unbiased and valuable insight into a providers' past track record, including its financial status, legal action taken against the organization and its commercial reputation. Certification schemes such as ISO27001 also provide customers with some assurances that a Cloud provider has taken certain steps in its management of information security risks.

V. DATA SECURITY [6] THREATS

There are several types of Data Security threats to which Cloud Computing is vulnerable:

A. Data Loss [7]

There are many ways to compromise data. Deletion or alteration of records without a backup of the original content is an obvious example. Unlinking a record from a larger context may render it unrecoverable, as can storage on unreliable media. Loss of an encoding key may result in effective destruction [8].

Solution for preventing the data [9] is to implement strong API access control; to encrypt and protect integrity of data in transit; to analyze data protection at both design and run time; to implement strong key generation, storage and management, and destruction practices; and to contractually specify provider backup and retention strategies

B. Data Integration

The integrity of data within complex Cloud hosting environments could provide a threat against data integrity [10]. A bad integration caused by incompatible interfaces or inconsistent policy enforcement may evoke both functional and non-functional impacts.

C. Data Stealing

This is the most traditional and common approach to breach a user account. The user account and password can be stolen by any means. As a result, the subsequent stealing of confidential data [11] can hamper the storage integrity and security of the Cloud.

Solution is “At the end of every session, the customer will send an e-Mail about the usage and duration with a special number to be used for log in next time”. In this way, the customer will be aware of the usage and charges as well as be availed with a unique number to be used every time to access the system. In Amazon EC2, a key pair is used to verify the authenticity of the customer.

D. Data combination and commingling

The Cloud Computing client needs to ensure whether its private data is stored separately from others or not. If they are combined or commingled with those of other clients’ data, then it is much more vulnerable or dangerous [12]. For example, viruses might be transmitted from one client to others. If another client is the victim of a hack attack, the attack might affect the availability or integrity of the data of other companies located in the same environment

VI. DATA PROTECTION AS A SERVICE

DPaaS [13] is a suite of security primitives offered by a Cloud platform, which enforces data security and privacy and offers evidence of privacy to data owners, even in the presence of potentially compromised or malicious applications.

To ensure a practical solution, the following goals relating to data protection as well as ease of development and maintenance were considered:

1) *Integrity:* The user’s stored data won’t be corrupted.
2) *Privacy:* Private data won’t be leaked to any unauthorized entity.
3) *Access transparency:* Logs will clearly indicate who or what accessed any data.
4) *Ease of verification:* Users will be able to easily verify what platform or application code is running, as well as whether the Cloud has strictly enforced their data’s privacy policies.
5) *Rich computation:* The platform will allow efficient, rich computations on sensitive user data.

A. Encryption: FDE versus FHE

Fully Disk Encryption (FDE) encrypts entire physical disks with a symmetric key, often in disk firmware, for simplicity and speed. With FDE [14], the keys reside with the Cloud platform, generally on or close to the physical drive: the Cloud application user isn’t involved in key management. Although FDE is effective in protecting private data in certain scenarios such as stolen laptops and backup tapes, the concern is that it can’t fulfill data protection goals in the Cloud, where physical theft isn’t the main threat.

Fully homomorphic encryption (FHE) [15] offers the promise of general computation on cipher texts. Basically, any function in plaintext can be transformed into an equivalent function in cipher text: the server does the real work, but it doesn’t know the data it’s computing. Naturally, this property gives strong privacy guarantees when computing on private data, but the question of its practicality for general Cloud applications still remains.

B. Architecture

Figure 2 illustrates example architecture for exploring the DPaaS design space. Here, each server contains a Trusted Platform Module (TPM) to provide secure and verifiable boot and dynamic root of trust.

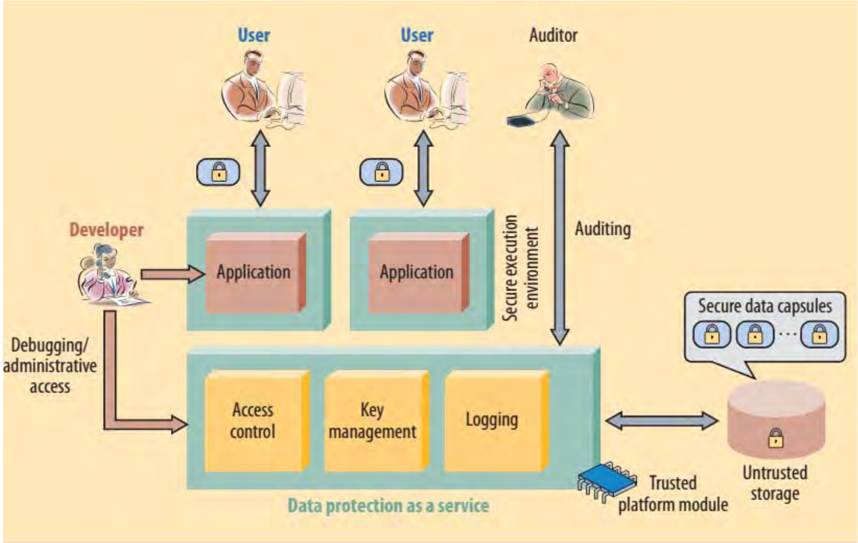

Fig 2. Architecture for DPaaS

A Secure Data Capsule (SDC) is an encrypted data unit packaged with its security policy. For example, an SDC might encompass a sharable document or a photo album along with its ACL. The platform can use confinement and information-flow controls to enforce capsules' ACLs.

To avoid unauthorized leakage of user data in the presence of potentially buggy or compromised applications, DPaaS confines the execution of applications to mutually isolated Secure Execution Environments (SEEs). Inter-SEE isolation has different levels, but stronger isolation generally exacts a greater performance cost due to context switching and data marshaling. At one end, a SEE could be a virtual machine with an output channel back to the requesting user. For performance reasons, it's possible to have a pool of VMs or containers in which the data state is reset before being loaded with a new data unit—similar to how a thread pool works in a traditional server.

The DPaaS approach places two additional requirements on the platform:

- It must be able to perform user authentication, or at least have a trusted way to know who's logged in and accessing the service; and
- It must rely on encryption and authenticated data store techniques to remove the need to trust the storage service.

C. Achieving data protection goals

DPaas uses a combination of encryption at rest, application confinement, information flow checking, and auditing to ensure the security and privacy of users' data. Application confinement isolates faults and compromises within each SEE, while information flow checking ensures that any information flowing among SEEs, data capsules, and users satisfies access-control policies. Controlling and auditing administrative accesses to data provides accountability. DPaaS can guarantee the integrity of the data at rest via cryptographic authentication of the data in storage and by auditing the application code at runtime.

Access controls, authorization, and auditing capability are common challenges for application developers. Incorporating these features within the platform is a significant improvement in terms of ease of use, and it doesn't constrain the types of computation that can be performed within a SEE. The platform logs common maintenance and batch processing tasks to provide accountability.

VII. DATA CENTER SECURITY

Data centers form the technical basis for Cloud Computing [16]. To this extent, it is important that every CSP ensures their systems are secure in compliance with the current state. This includes permanent monitoring of access, for example using video monitoring systems, movement sensors, alarm systems and trained security personnel.

Modern fire protection precautions also need to be taken, and tested on a regular basis. The data centers should be located far enough away from each other geographically so that a controllable damage event, e.g. fire, explosion, road, rail, water or air accidents and natural disasters with a limited impact such as flooding does not simultaneously affect both the data center originally being used and the one containing the backup capacities.

VIII. NIST GUIDELINES ON SECURITY AND PRIVACY IN PUBLIC CLOUD COMPUTING

Since the data stored in a public Cloud typically resides in a shared environment collocated with data from other customers, the NIST report strongly recommends that access to the data should be controlled and the data should be kept secured [17]. These requirements are also applicable for the data that is migrated within or between Clouds. In addition, data can take many forms in the Cloud. For example, for Cloud-based application development, data may include the application programs, scripts, and configuration settings, along with the development tools.

The NIST report recommends two methods for keeping data away from unauthorized users: (i) access controls, and (ii) encryption. Access controls are typically identity-based, which makes authentication of the user's identity an important issue in Cloud Computing. However, lacking physical control over the storage of information, encryption is the only way to ensure that it is truly protected. In addition, data must be secured while at rest, in transit, and in use, and access to the data must be controlled.

The NIST report observes that the security of a system that employs cryptography depends on the proper control of central keys and key management component. Currently, the responsibility for cryptographic key management falls mainly on the Cloud consumer. Key generation and storage is usually performed outside the Cloud using hardware security modules, which do not scale well to the Cloud paradigm. NIST also recommends that before proceeding in Cloud environments where the Cloud provider provides facilities for key management, the organization must fully understand and weigh the risks involved in the processes defined by the Cloud provider for the key management lifecycle. Hence, the cryptographic operations performed in the Cloud become part of the key management process and, therefore should be managed and audited by the organization. In a public Cloud, data from one consumer is physically collocated (e.g., in an IaaS data store) or commingled (e.g., in a SaaS database) with the data of other consumers, which can complicate matters. Hence, NIST recommends that sufficient measures should be taken to ensure that data sanitization should be performed appropriately throughout the system lifecycle.

IX. CONCLUSION

In today's global competitive market, companies must innovate and get the most from its resources to succeed. Cloud computing helps IT enterprises use various techniques to optimize and secure application performance in a cost-effective manner. Cloud Computing is a relatively new concept that presents a good number of benefits for its users; however, it also raises some security problems which may slow down its use. 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 service delivery models and in all stages of data life cycle.

The massive concentrations of resources and data present a more attractive target to attackers, but Cloud-based defenses can be more robust, scalable and cost-effective than traditional ones. To help reduce the threat, Cloud Computing stakeholders should invest in implementing security measures to ensure that the data is being kept secure and private throughout its lifecycle.

X. FUTURE WORK

For data security and privacy protection issues, the fundamental challenges are separation of sensitive data and access control. So, 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. It is strategic to develop new mechanisms that provide the required security level by isolating virtual machines and the associated resources while following best practices in terms of legal regulations and compliance to SLAs. Among other requirements, such solutions should employ virtual machine identification, provide an adequate separation of dedicated resources combined with a constant observation of shared ones, and examine any attempt of exploiting cross-VM and data leakage. A deeper study on current security solutions to manage Cloud Computing virtual machines inside the Cloud providers should be a focus of future work in this area.

REFERENCES

[1] Zhao.G, Liu.J, Tang.Y (2009), Cloud Computing: A Statistics Aspect of Users. , First International Conference on Cloud Computing (CloudCom), pp 347–358

[2] Zhang.S, Chen.X (2010), Cloud Computing Research and Development Trend , Second International Conference on Future Networks (ICFN'10), IEEE Computer Society, pp 93–97

[3] Ogigau.F, Data security and confidentiality issues, Journal of Defence Resources Management, Vol.3, Issue 2(5),2012.

[4] Deyan.C, Data Security and Privacy Protection Issues in Cloud Computing, International Conference on Computer Science and Electronics Engineering

[5] Jaydip.S, Privacy Issues in Cloud Computing, Innovation Labs, Tata Consultancy Services Ltd., Kolkata, INDIA)

[6] Security Guidance For Critical Areas of Focus in Cloud ComputingV2.1, https://Cloudsecurityalliance.org/guidance/csaguide.v2.1.pdf

[7] Kangchan.L Security Threats in Cloud Computing Environments, International Journal of Security and Its Applications Vol. 6, No. 4, October, 2012

[8] Web Source accessed from http://en.wikipedia.org/wiki/Microsoft_data_loss_2009

[9] Top Threats to Cloud Computing V1.0, accessed from <https://Cloudsecurityalliance.org/topthreats/csathreats.v1.0.pdf>.

[10] Security Guidance for Critical Areas of Focus in Cloud Computing V2.1, accessed from <https://Cloudsecurityalliance.org/research/security-guidance/>, (2011) November

[11] Zunnurhain.K, Susan.V, Security Attacks and Solutions in Clouds

[12] Ashktorab.V, Security Threats and Countermeasures in Cloud Computing, International Journal of Application or Innovation in Engineering & Management (IJAIEM)

[13] Dawn.S, Elaine.S, Fischer.I, Cloud Data Protection for the Masses

[14] Maniatis.P, Do You Know Where Your Data Are? Secure Data Capsules for Deployable Data Protection, 13th Usenix Conf. Hot Topics in Operating Systems, Usenix, 2011

[15] Gentry.C, Fully Homomorphic Encryption Using Ideal Lattices, Proc. 41st Ann. ACM Symp. Theory Computing (STOC 09), ACM, 2009, pp. 169-178.

[16] Michael.H, Security Recommendations for Cloud Computing Providers.

[17] Wayne.J, Timothy.G, Guidelines on Security and Privacy in Public Cloud Computing, NIST, Draft Special Publication 800-144, January 2011.