Securing Cloud Data in Transit using Data Masking Technique in Cloud Enabled Multi Tenant Software Service

S. Selvakumar* and M. Mohanapriya

Department of CSE, Karpagam University, Coimbatore - 641021, Tamil Nadu, India;
selva_harur@yahoo.co.in, mohanapriyaasaithambi@gmail.com

Abstract

Background/Objectives: The paper discusses about the data security issues in the cloud computing environment. It employs data masking to hide sensitive data from cloud services thereby ensuring reliability and trust in the cloud environment. Methods/Statistical Analysis: Data access in the cloud environment can be categorized into three such as at rest, at transit, in use. The main aim of this paper is to integrate security in data masking techniques. We employ the existing mechanisms to the cloud environment to secure the data with virtual machine masking and platform masking. Findings: The masked data is transmitted to the processing environment. The services in cloud utilize this masked data for processing. It is comparatively secured when compared to the conventional technique. This mechanism increases the trustworthiness and can be masked dynamically or statistically in application or database based service environments. Application/Improvements: The main application of this research is to serve people with secured cloud, thereby overcoming the data security issues.

Key words: Cloud Environment, Cloud Service Provider, Data Masking

1. Introduction

All the business domains started utilizing cloud computing for their consumption or as a platform to deliver their services. As a result of this new business models were introduced and new services in cloud delivery model were introduced. The cloud services are also implemented in number of ways on different deployment models. All these cloud services employ data at different level. As the number of cloud services, the number of users utilizing the cloud services also increases, thereby increasing the volume of data generated. The data in the cloud environment are required by the cloud services and in turn data are generated by the cloud services. The data in the cloud may be under the logical control of the user but the data physically reside in the infrastructure of cloud service provider.

Maintenance and control of the data in cloud is tedious task. The users should have trust on the cloud service providers that their data in cloud is free from breach and the services satisfy the compliance and regulatory measures. Although the cloud service life cycle and the resources are maintained by the cloud service providers, users do not trust generally the service providers to provide them data access due to the trust and security considerations. Moving data into the cloud may cause many hardware complexities. The internet based hardware services provide various ways in the these cloud services can be implemented.

The security considerations must be addressed to make the cloud sustainable. This can be achieved by including real time security intelligence, securing the data, preventing threats and attacks. In a cloud environment, there are multiple complex data security challenges to be addressed which include the need to protect confidential data, adhering to regulations. There are still lack of standards persist in the cloud environment. In a cloud environment offering services through multi-tenant mode where the same instance is shared by multiple users, the same infrastructure platform and the

* Author for correspondence
application along with the data base, high priority is to be provided for the data security. The complexities increase in case of interoperable cloud environment with mixed trust levels. The cloud service provider should plan and be ready to face the additional risks concerned with security challenges.

Most of the cloud service providers have integrated various encryption mechanisms for data security in their cloud environment; the focus is less on in management of the keys, access control and data access monitoring. In case the keys for encryption are not protected, vulnerability by hackers increased the theft which also extends to access control. Any cloud environment offering services with data centric approach should focus on strong encryption, strong access control, management of keys and security intelligence which can protect the data and can provide the required level of security. The required security level can be implemented in layered approach. An effective cloud solution requires the following capabilities such as data lockdown, access policies and security intelligence. The first level focus on the data and ensures that the data is not readable and the system provide required key management, then the focus in on access policies which ensures that the privileged users cannot access the primitive data. The third level is to include security intelligence.

Our focus is to address the data security issues in the cloud environment. Data access in the cloud environment can be categorized into three such as data at rest, data at transit and data in use. Data at rest represents the data stored in the cloud environment, data in transit represents that the movement of data from storage to virtual machines and within the infrastructure, platform, software services, between clouds in interoperable cloud environment. Data in use represents the data when accessed by the services in cloud environment.

Our focus in the paper is to secure the data at transit and data in use by integrating the data masking technique. Data masking is a technique to hide the sensitive data from the users. In case of the cloud environment, our approach employs data masking to hide sensitive data from cloud services thereby ensuring reliability and trust in the cloud environment. Our approach of data masking for data at transit and data in use enhances the existing security measures in a conventional cloud environment.

The paper is organized in the following fashion; first we introduce various security measures in cloud environment followed by various existing works focusing on security in cloud environment. Then we explain our approach with the implementation and results. Finally we conclude by providing the concluding remarks and future works that can be carried out of this work.

2. Related Work

Multi-prime RSA algorithm encrypts the data before it is stored in the cloud environment. After successful encryption, the data is stored into the cloud. The encrypted data is decrypted when an authorized person request for data. In spite of the benefits there are lot of issues prevails such as multi tenancy, key management. To ensure that the data integrity to be managed we use a third party auditor that help the provider's environment alone. The data has been prevented using RSA algorithm. Whenever the user wants to store their data in cloud, the user can login into the server and encrypt the desired data. This RSA algorithm is performed using the set of prime numbers. The data is automatically sent to the cloud environment. The private key is kept confidential. Whenever the user wants to view the data the private key is provided by the user to the cloud and the data is so retrieved. The authors express that the data is backed up using the cloud services. Firewall helps in preventing the data from the unauthorized attacks.

Firewalls are implemented to prevent information which resides in the cloud from DDoS and Dos attacks. It express that data should be backed up as well as appropriate level of data mirroring is implemented.

The security issues occurs when the external users use the data in the unauthorized way. The data stored in the cloud can be viewed by the user. The authorization and the authentication can be done using the id and password. To protect the host operating system from the attacker, the minimal operating system is chosen to be stripped for all unnecessary service.

Symmetric and asymmetric key encryption methods help in encrypting the data in a very preventive manner so that no third party can access the data without the authentication of the personal user. This makes the system very effective and a secured one.

A secure cloud storage system supporting privacy-preserving public auditing and further extend our result to enable the TPA to perform audits for multiple users simultaneously and efficiently. By integrating the HLA with random masking, our protocol guarantees that the TPA could not learn any knowledge about the data content stored in the cloud server during the efficient auditing.
process. The aggregation and algebraic properties of the authenticator further benefit our design for the batch auditing.

An efficient framework is proposed for privacy preservation. Considering the information loss metrics, one-dimensional quasi identifiers are focused and studied the properties of optimal solutions for k-anonymity. To generalize, an efficient multi-dimensional quasi identifiers using space mapping methods are proposed. A framework is developed for solving the k-anonymity problems, by mapping the quasi-identifiers to one dimension. However, a set of properties have been identified for the optimal 1-D solution. Guided by these properties, efficient algorithms are developed at the 1-D space. Popular transformations namely the Hilbert curve is used to solve the anonymity problem. the projected algorithms are linear to the input size. Therefore they are applicable to very large datasets.

3. Data Masking

In this section, we provide the architecture of integrating data masking technique. Data masking is the process of replacing the sensitive information with virtual information or duplicated information such that it is not misused by any intruder. There are various techniques for data masking such as substitution, shuffling, Encryption, Nulling out / Truncating, Masking out data, Table to table synchronization, Synchronizing between different data types, Cross database synchronisation, cross server synchronisation, cross platform server synchronization, selective masking, flat file masking, nulling, blurring. We employ the existing mechanisms to the cloud environment to secure the data with virtual machine masking, platform masking.

Although various advanced encryption techniques exist which can encrypt the data, these encrypted data are applicable only to data at rest or to data in transit. In the situation where data is in use, the encrypted data cannot be directed processed by the cloud service. Hence the data is to be decrypted and data in raw format is to be utilized by the services. If there is a flaw in the application or in services, the chances of data being hacked are high. In case of a cloud based multi-tenant service, the chances of compromising data when data in use is extremely high.

3.1 Architecture

Our approach masks the data before the data is in transit. The masked data is transmitted to the processing environment. The services in cloud utilize this masked data for processing. Our approach is comparatively secured when compared to the conventional technique. This mechanism increases the trust worthiness of the cloud service provider. The data in cloud environment can be masked dynamically or statistically in application or data base based service environments. This data masking technique is considered as one of the best technique to desensitize data.

Data masking techniques involves mainly of three methods they are

- Synchronisation and synthetization of the symbolic data
- Anonymization of the data
- Masking and tokenization of the present data there by providing security also called as pseudonymization

There are five rules for data masking [15] such
Securing Cloud Data in Transit using Data Masking Technique in Cloud Enabled Multi Tenant Software Service

as masking must not be reversible; results must be representative of source data; referential integrity must be maintained; only mask non-sensitive data if it can be used to recreate sensitive data; masking must be repeatable process.

The architecture of the cloud based data masking technique is provided in Figure 1,

The interoperable cloud environment is considered in our architecture. A cloud environment consists of Cloud head controlling the entire cloud environment, followed by its cluster head, the underlying nodes. These nodes host virtual machines. In a distributed multi-tenant application service, one or more application service instance and one or more data base instance is employed to deliver the desired service. This integrated service delivery on multi-tenant mode is achieved through orchestration. The masking module is introduced in the architecture that masks the data and services for enhanced security. The masker module plays a vital in orchestration of services and in case of migration from one cloud to other cloud service and also in case of interoperability between clouds or interoperability within the cloud. In a cloud based n-tier architecture, the application service and database service are isolated. The architecture makes the database service and application service non interdependent. Multiple application service can be integrated to deliver the specified multi-tenanted software as a service. This exhaustively shared mode arises a need for complete level of security through masker module.

The data masker module components are explained in the Figure 2.

The masker component consists of the mentioned functional domains such as meta data which takes up the meta data of the source data to be masked for processing, dynamic masking engine which is responsible for masking the required data dynamically as the services are getting provisioned dynamically in the cloud, software firewall to secure the masker component with the configured security policies, masking rules domain where the rules for masking are configured, cloning functionality is responsible for cloning the data, security audit functionality is responsible for auditing the entire functionality of the masker, data validation module validates the masked data with required stamps, data gathering module is responsible for streaming, manager is the complex module comprising of query execution manager, data source manager, security manager and workflow manager. The processing is split up into three functionality such as extract, convert and insert / load processing.

The constraints involved in masking are format preservation, data type preservation, gender preservation, semantic integrity, referential integrity, aggregate value, frequency distribution and uniqueness. Although there are two types of masking such as proxy based masking and dynamic masking, our focus in on dynamic masking on a distributed mode as this implies on dynamic deployment property of cloud.

The algorithm used for the masking functionality is described in the following section. For masking first we employ statistical distribution. Statistical distribution is generally a description of data masking techniques.

A distribution is generally represented as the measure and distribution of data

\[ P_x(s) = \frac{1}{P(S)} \int_P x(s) \, dP(s) \]

Let `x` is the measure \( P_x \) on \( S \) defined by setting

\[ P_x(A') = P\{s \in S : x(s) \in A'\} \]

where \((S,\mathcal{S},P)\) is a probability space, \((S,\mathcal{S})\) is a measurable space, \( P \) a measure on \( \mathcal{S} \) with \( P(S)=1 \).

The average distribution function prob. \( P \) determines as usual an average distribution function \( F \) according to the relation \( F(X) = \int G(x) \, dP(G) \). Consider the mapping \( T \) of \( A \) into \( A \) defined by

\[ F_{\mu}(x) = \int_{A} G(x) \, dP(G) \]

Consider the mapping \( T_{\mu} \) of \( \Delta \) into \( \Delta \) defined by

\[
\begin{align*}
(T_{\mu}F)(x) &= \int_{0}^{1} \int_{0}^{1} \beta F\left(\frac{x}{a}\right) \mu(\alpha d\alpha, d\beta) + \\
&\int_{0}^{1} \int_{0}^{1} (1 - \beta) F\left(\frac{x - \alpha}{1 - \alpha}\right) \mu(\alpha d\alpha, d\beta).
\end{align*}
\]
As the next stage we employ, Substitution method for recurrence relations, which is a way to compute the asymptotic complexity and plug it directly into the recurrence relation. We plug the expression \( kn \lg n \) into the merge-sort recurrence relation:

\[
kn \lg n \geq 2k(n/2) \lg (n/2) + c_n
\]

\[
= kn \frac{(n/2) \lg n}{n} + c_n
\]

\[
= kn (\lg n - 1) + c_n
\]

\[
= kn \lg n - kn + c_n
\]

\[
= kn \lg n + (c_n - k) n
\]

\( k \) makes inequality come true for sufficiently large \( n \), it holds if \( k \geq c_n \). Therefore the function involved is \( O(n \lg n) \).

4. Implementation Results and Discussion

The masking component is implemented over cloud environment in the following fashion. The entire cloud environment is implemented on open source software. Eucalyptus is used as the cloud middle ware and management tools and Xen as a virtualization manager. One cloud environment is implemented with two clusters. Each cluster have 2 nodes capable of hosting 2 virtual machines each. Cloud enabled distributed applications is deployed on the environment with n tier architecture. Two sets of application service is deployed on one VM and database service on other VM. This application service and database service are orchestrated to offer a cloud enabled service. We introduce the masking component into the structure. To exploit the interoperability scenario, we simulate the cloud environment with cloudsim enabled service. We introduce the masking component into it. We experimented the masking and found it to be encouraging to our approach.

5. Conclusion and Future Enhancement

This paper helps in securing the data at transit and data in use by integrating the data masking technique. In a distributed cloud environment, our approach employs data masking to hide sensitive data from cloud services thereby ensuring reliability and trust in the cloud environment. Our approach of data masking for data at transit and data in use enhances the existing security measures in a conventional cloud environment. This is employed to a larger extent in a distributed multi-tenant application. This multi-tenant scenario involves a n-tier architecture where multiple application service and database service are orchestrated for a software service.

Our approach masks the data before the data is in transit. The masked data is transmitted to the processing environment. The services in cloud utilize this masked data for processing. Our approach is comparatively secured when compared to the conventional technique. This mechanism increases the trust worthy of the cloud service provider. The data so present in the cloud environment can be masked dynamically or statistically in application or data base based service environments. This data masking technique is considered as one of the best technique to desensitize data. Few of the other works that can be extended are the current focus is on data in transit and data in use, where as same can be extended to data in rest. Also the current focus is only to the software services database whereas this can be extended to the virtual machine data for platform service.

6. References