Confidentiality of Database for Secure Information in Cloud Computing Using Table Collide Storage Model

Dr. K. Thilagam¹ S. Lakshmi Priya²
¹ Associate Professor, Department of Computer Application, Karpagam Academy of Higher Educations, Coimbatore
² Research Scholar, Department of Computer Science, Karpagam Academy of Higher Educations, Coimbatore

Abstract: Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources. Cloud computing provides computation, software, data access, and storage services that do not require end-user knowledge of the physical location and configuration of the system that delivers the services. Since the data transmission on the internet or over any networks is vulnerable to attack, privacy preservation and data integrity have become two of the most critical security issues related to user data. In conventional paradigm, the organizations have the physical possession of their data and hence have an ease of implementing better data security policies. But in case of cloud computing, the data is stored on an autonomous business party, who provides data storage as a subscription service. The users have to trust the cloud service provider (CSP) for security of their data. Cloud database can be decomposed and placed in different service providers’ servers at cloud computing to secure data from any intruder. Here the challenge lies in security and also placing data at different locations including how, when and where to place in cloud storage to maintain confidentiality. This article describes tables collide storage model with private key table structure, which ensures confidentiality in cloud computing by splitting table into small pieces and implicitly performs encryption through collide different table’s pieces before storing in cloud environment.

1. INTRODUCTION

With growing dependence on computers and eventually the Internet, the need for databases and registries, protection of an individual’s privacy is now one of the greatest challenges in research.

Cloud computing technology includes many technologies such as the autonomic computing virtualization, utility computing, service oriented architecture and many others. The purpose of these technologies is to provide scalable, shared resources, software and hardware services over the network. The cloud, termed ‘as a service’, is referred to as providing something as a service over the network. There are 3 types of services provided by cloud, they are: Software as a Service (SaaS), Platform as a Service (PaaS), Infrastructure as a Service (IaaS) and many more[19][24]. All the provided services are based on the policy of on-demand fashion in which users can pay only for their required usage. Today, many cloud service providers such as Amazon’s EC2 and S3, Microsoft’s Windows Azure Google’s App Engine are providing the facility to different users. Users who cannot afford such a huge cost build their own infrastructure, so they can have their work done by the help of cloud providers at a minimum cost. Depending on the type of users and the hosting environment the cloud architecture can be divided into four types:

1. Public Cloud
2. Private Cloud
3. Hybrid Cloud
4. Community Cloud

Public cloud provides services, which are hosted for public usage and anyone can have their data stored and services get done using this cloud. Data security is the most important issue here
Here resources are dynamically provisioned on a fine-grained, self-serviced basis over the Internet. They are less secure than the other models.

In the private cloud, the data access and service usage are restricted to one authorized user only. The private cloud is set up within an organization’s internal enterprise data center. In the private cloud, scalable resources and virtual applications provided by the cloud vendor are pooled together and available for cloud users to share and use.

In the hybrid cloud, it shared by a limited number of organizations and it combined the features of both the private and public clouds. The hybrid cloud is a private cloud linked to one or more external cloud services, centrally managed, provisioned as a single unit, and circumscribed by a secure network.

This cloud also provides the service called database-as-a-service (DaaS). In this the data owner can outsource his data/files on cloud server for reducing space cost as well as maintenance cost and only the authorized user can query/request this data. Here issues of confidentiality include: data owner mostly concerned on the security from unauthorized access, integrity means correctness of the file/data after outsourcing to the proxy server and should not be modified by unauthorized user or even though by the proxy server. So, this is the reason why it has become a major research problem among the research community and it is turning more serious day by day.

2. REVIEW OF LITERATURE

2.1. Cloud Computing and its Security

Gartner defines Cloud computing as massively scalable IT-enabled capabilities which are delivered as a service to external customers using internet technologies. Google’s Google App Engine and Amazon’s EC2 service are examples of Cloud computing. Cloud computing has unique attributes that require risk assessment in areas such as data integrity, recovery, and privacy. Gartner has also identified seven security issues that need to be addressed before enterprises consider switching to the cloud computing model.
According to Mather, Tim et al.[17], Cloud computing is a broad and diverse phenomenon. Much of its growth represents a transfer of traditional IT services to the new cloud model, but there is also scope for creation of substantial new businesses and revenue streams. Cloud computing offers the promise of more efficient and cost-effective computing to facilitate information use, but also expands known risks and introduces new risks yet to be discovered and managed.

Qian, Ling et al. [18] state that the cloud computing has been evolved from internal IT system to public service, from cost-saving tools to revenue generator, and from ISP to telecom. The authors have analyzed the concept, history, pros and cons of cloud computing as well as the value chain and standardization effort.

Zhao et al. [19], in their view, explain that cloud computing has numerous benefits for its users and at the same time, it raises some security problems which slow down its use. They have presented security issues for cloud models: IaaS, PaaS, and IaaS and pointed out that storage, virtualization, and networks are the biggest concerns in cloud computing.

In cloud security alliance[24], some useful technologies and best practices for securing data within various modules include Infrastructure as a Service (IaaS), Platform as a service (PaaS) and Software as a Service (SaaS). A survey by Cloud Security Alliance (CSA) and IEEE indicates that enterprises across sectors are eager to adopt cloud computing but that security is needed both to accelerate cloud adoption on a wide scale and to respond to regulatory drivers. It also details that cloud computing is shaping the future of IT but the absence of a compliance environment is having a deep impact on cloud computing growth.

The Cloud Computing Use Case Discussion [21] Group discusses the different Use Case scenarios and related requirements that may exist in the cloud model. They consider use cases from different perspectives including those of the customers, the developers and security the engineers.

ENISA[22] investigated the different security risks related to adopting cloud computing along with the affected assets, the risks likelihood impacts, and vulnerabilities in the cloud computing which may lead to such risks.

Dahbur et al. [23] describe the desirable benefits of cloud, risks and security concerns that must be considered and addressed correctly. They categorize many of the security issues introduced by the "cloud" survey the risks, threats and vulnerabilities, and make the necessary recommendations that can help to promote the benefits and mitigate the risks associated with cloud computing.

Sasirekha and Hemalatha [30,31] observe that cryptographic techniques are found to be very efficient in dealing with a number of software threats and attacks. They have identified that code encryption has received much attention in the field of software security. Number theoretic and hadamard matrix methods have been used for transferring the code to the clients in a secured way.

According to Hashizume et al. [25], cloud computing appears as distributed services with all resources as services and delivered over the internet. It enriches accessibility on demand in an efficient and more cost effective manner. Challenges in security leads to compliance, privacy and legal issues. There is an uncertainty of moving security applications as a cloud. This uncertainty has created great attraction for research on security in cloud computing. The cloud solution provider must ensure privacy and security control over their applications and data.

Garg et al. [26] in their vision, they propose to extend control measures from the enterprise to the cloud through the use of Trusted Computing and applied cryptographic techniques. These measures should alleviate much of today’s fear of cloud computing.
2.2. Security through cryptography

Juels and Kaliski [27] describe a proof of retrievability model. It ensures possession of data on archive service system by adding ‘sentinels’ blocks for detection purpose, and file encrypted to protect the special blocks.

According to Paul,A.J. et al. [28], substitution and diffusion operations, based on the matrix, facilitate fast conversion of plain text and images into cipher text and cipher images. They have proved that their algorithm is much faster than AES [28].

Paul,A.J. et al. [29] have developed an encryption algorithm with substitution mapping, translation and transposing operations. This has two advantages over traditional algorithms. They are simpler and much faster and highly secured.

Haitner et al. [32] proved that two-to-one hash function achieves better parameters compared with one-way hash function.

Ren et al.[33] proposed a dual protection with the DES algorithm and RSA algorithm securing data transmission in the Bluetooth system in an efficient manner.

Vellaikannan et al. [15] have proposed a novel method for sharing messages confidentially. This method is applicable to lengthy messages. Diagonal matrices induced from Quadratic forms are preferred for encoding as their inverses can be easily obtained. This technique provides a transaction of the least amount of messaging between the sender and the receiver. (It is sufficient to know the codes of use and the Quadratic form). Here the security is assured. Higher order diagonal matrices are preferred as their inverses can be easily found. When the size of the message is too large, new string operations may be defined and the message can be split for suitable process.

2.3. Database Level Encryption

Database Level Encryption (DLE) secures data as it is written to and read from a database. The encryption is applied to the Db at various granularities, such as database, tables, columns (most frequently), and rows. It can be related with some logical conditions for selecting affected data, too. Several database encryption schemes have been proposed in the literature by the authors Pagano, Francesco(2015) and Shmueli, Erez, et al.(2010) [45][46].

According to Pagano, Francesco (2015) and Davida (1981) a scheme based on the Chinese-Remainder theorem is proposed, where each row is encrypted using different sub-keys for different cells. This scheme enables encryption at the level of rows and decryption at the level of cells [45][47].

An extension of the previous scheme that supports multilayer access control is proposed by Pagano, Francesco (2015) and Hwang (1997). Their work classifies subjects and objects into distinct security classes that are ordered in a hierarchy in which an object with a particular security class can be accessed only by subjects in the same or a higher security class [45][50].

Buehrer and Chang [48] proposes encryption for a database based on Newton's interpolating polynomials.

Chang et al. [49] is focused on the RSA public-key scheme and suggests two database encryption schemes: one column oriented and the other row oriented.

The SPDE scheme is presented by Hwang (1997). This scheme encrypts each cell in the database individually together with its cell coordinates (table name, column name and row-id) to obtain different cipher-text values for equal plaintext values (against static analysis), and to prevent a tuple being moved to a different location (against splicing attacks) [50]. Kühn and Ulrich (2006) performed the analysis of database and index encryption scheme problems and their fixes[51].
Authors Pagano, Francesco(2015) and Shmueli, Erez, et al(2010) referred to disadvantages of the DLE schemes. One crucial aspect is that all but the last these schemes is that the basic element in the database is a row and not a cell. Thus the structure of the database is modified. In addition, all of those schemes require re-encrypting the entire row when a cell value is modified. Thus, in order to perform an update operation, all the encryption keys should be available [45][46].

2.4. Binding Substitution, Translation & Transposing

Paul et al. have developed an encryption algorithm with substitution mapping, translation and transposing operations which has two advantages over traditional algorithms. They are simpler and much faster and highly secured. This algorithm is suitable for high speed encryption applications. Their article concludes that combination of poly alphabetic substitution, translation and transposition makes the decryption extremely difficult without having the secret key.

2.5. Percentage Level Of Security Challenge

Chart from cloud report shows that the percentage level of security challenges decreases when organizations gain expertise with cloud, then the responsibility of security shared by both cloud service providers and enterprise and this responsibility differs depends on cloud service model. The details are shown in figure 3. Google Docs [53] says that during June 2011, attack on Google from Jinan and China shows that on-line document storage in cloud environment gives less security to the organizations.

2.6. Redundant Array Of Independent Net-Storages (RAIN)

Jaatun et al. [52] focused on security in cloud storage. In their related study, they found out that the approach of Singh et al. did not focus on how to split the data to be stored and RACS not offer privacy or confidentiality. Jaatun et al. described about traditional botnets (collection of compromised computers), where C&C (Command & Control) botnet is created by infected PCs. The infected host obeys without verifying the identity of the issuer. Jaatun et al. suggested new types of cloud service providers, namely pure cloud solution and intelligent client model. The pure cloud solution keeps all processing in the cloud and leaves all information with the service provider with trust. Both processing and storage providers can access only pieces of encrypted data which guarantees secrecy. The intelligent client model deals information at the C&C side itself instead of under a hopeless service provider.

The RAIN approach split the data into segments and distributes them among multiple service providers for storage. The relationship between segments kept as secret at the C&C side to re-assemble segments. Small segments ensure the confidentiality in cloud computing. Various protocols were described for splitting data, processing data in cloud is for re-assembling.
Here partitioned data stored based on the amount of free space in the cloud. If there is more space in the cloud, then this protocol may continue to store attributes and this may leave the data unsecured in cloud.

3. CLOUD COMPUTING

3.1. Cloud Storage Characteristics

The following table presents the list of some of the characteristics of cloud storage and its description [5][6][38].

<table>
<thead>
<tr>
<th>S. No</th>
<th>Characteristics</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Manageability</td>
<td>The ability to manage a system with minimal resources</td>
</tr>
<tr>
<td>2</td>
<td>Access Method</td>
<td>Protocol through which cloud storage is exposed</td>
</tr>
<tr>
<td>3</td>
<td>Performance</td>
<td>Performance is measured by bandwidth and latency</td>
</tr>
<tr>
<td>4</td>
<td>Multi-tenancy</td>
<td>Support for multiple users or tenants</td>
</tr>
<tr>
<td>5</td>
<td>Scalability</td>
<td>Ability to scale up to meet higher demands or load in a graceful manner</td>
</tr>
<tr>
<td>6</td>
<td>Data availability</td>
<td>Measure of a system’s uptime for data availability</td>
</tr>
<tr>
<td>7</td>
<td>Control</td>
<td>Ability to control a system, in particular to configure for cost, performance, or other characteristics</td>
</tr>
<tr>
<td>8</td>
<td>Storage efficiency</td>
<td>Measure of how efficiently the raw storage is used</td>
</tr>
<tr>
<td>9</td>
<td>Cost</td>
<td>Measure of the cost of the storage</td>
</tr>
<tr>
<td>10</td>
<td>Location independence</td>
<td>The customer has no control or knowledge over the exact location of provided</td>
</tr>
<tr>
<td>11</td>
<td>Flexibility/Elasticity</td>
<td>Users can rapidly access computing resources, as required, without human interface.</td>
</tr>
<tr>
<td>12</td>
<td>Reliability</td>
<td>Provides the complete use of multiple redundant sites, which makes cloud computing suitable for business continuity and disaster recovery.</td>
</tr>
<tr>
<td>13</td>
<td>Sustainability</td>
<td>This can be achieved by means of improved resource utilization, more efficient systems, and carbon neutrality</td>
</tr>
</tbody>
</table>

3.2. Privacy Vs Confidentiality

The advantage of cloud storage is cost saving. The prime disadvantage is security. Since the security is not provided in cloud storage, many companies adopt their unique security structure. For example Amazon has its own security structure. Since the data placed in cloud storage is accessible to everyone, security is not guaranteed [38].

Privacy is defined in terms of a person having control over the extent timing, and circumstances of sharing any user’s data either physically, behaviourally or intellectually with others. Privacy refers to the right of individuals to limit access by others to aspects of their person [34].

Confidentiality on the other hand is the process of protecting an individual’s privacy. It pertains to the treatment of information that an individual has disclosed in a relationship of trust, with the expectation that this information will not be divulged to others without proper permission [35]. The need to keep personal information private is often weighed against the need to share personal...
information that has the potential to benefit the public good. Sharing information from DNA sequencing, databanks and repositories, quality-assurance efforts, and public health measures is essential for the development of new medical treatments. It is important for service providers to understand how these competing values of maintaining confidentiality and how to balance the cloud providers’ goals against these competing needs.

Privacy relates to the research participant’s direct disclosure to the researcher; confidentiality relates to the extent to which the cloud security providers protect the user’s private information.

Benefits of Maintaining Confidentiality:

- It helps establish trust between the cloud security provider and the cloud user.
- It reduces worry on the part of the individual user.
- It safeguards the cloud user’s dignity.
- The user feels respected.
- It gives the cloud user a mode of control and promotes autonomy.

No individual should risk harm due to disclosure of his/her private information as a result of his/her database being maintained in the cloud [36].

3.3. Cryptography In Databases

Confidentiality, integrity and availability are the main properties of database protection [44]. Confidentiality has been defined by the International Organization for Standardization (ISO) in ISO-1779910 as "ensuring that information is accessible only to those authorized to have access"; data integrity assures that none can modify the information without a trace; availability provides access to data by authorized users within a reasonable time. Along the years, a lot of ACP (Access Control Policy) has been defined, based on database model (relational rather than object) and policy control [45].

Traditionally, ACPs are based on the assumption that the DBA (DataBase Administrator) is trusted, but this assumption no longer holds in outsourced data centres and in the Cloud, where the platform-as-a-service (PaaS) provider is external to data owner. A solution to this problem is that the DBMS treats only raw-data, encrypted in such a way that DBA (or another intruder) cannot read the information. There are three main categories of database encryption [45]: storage level encryption, database level encryption, and application level encryption.

3.4. Securing Data In Cloud

There are three key areas of concern related to security and privacy of data [38]:

- Location of your data
- Control of your data
- Secure transfer of your data

A. DATA LOCATION IN CLOUD:

1. Data transfer across country borders: A global company with subsidiaries or partners (or clients for that matter) in other countries may be concerned about cross-border transfer of data due to local laws. Virtualization makes this an especially tough problem because the cloud provider might not know where the data is at any particular moment.

2. Co-mingling of data: Even if the data is in a country that has laws then the user is comfortable with the data which may be physically stored in a database along with the data from other companies. This raises concerns about virus attacks or hackers trying to get at another company’s data.

3. Secondary data use: In public cloud situations, the data or metadata may be vulnerable to alternative or secondary uses by the cloud service provider. Without proper controls or service level agreements, the user data may be used for marketing purposes and merged with data.
from other organizations for other alternative uses also. The recent uproar about Facebook mining data from its network is an example. The service provider may own any metadata is created to help and manage the data, minimizing the user’s ability to maintain control over the data [38].

B. DATA CONTROL IN CLOUD:

Data controls include the governance policies set in place to make sure that the user data can be trusted. The integrity, reliability, and confidentiality of the data must be beyond reproach. For example, assume that any user is using a cloud service for word processing. The documents that are created are stored with the cloud provider. These documents belong to the user’s company and the user has the control access to those documents. No one should be able to get them without the owner’s permission, but perhaps a software bug lets other users to access the documents. This privacy violation results from a malfunctioning access control. Here is a sampling of the different types of controls designed to ensure the completeness and accuracy of data input, output, and processing: Input validation controls to ensure that all data input to any system or application are complete, accurate, and reasonable. Processing controls ensure that data are processed completely and accurately in an application. File controls make sure that data are manipulated accurately in any type of file (structured and unstructured). Output reconciliation controls ensure that data can be reconciled from input to output. Access controls to ensure that only those who are authorized to access the data can do so. Sensitive data must also be protected in storage and transfer [38][40].

C. SECURING DATA FOR TRANSPORT IN CLOUD:

A virtual private network (VPN) is one way to manage the security of data during its transport in a cloud environment. A VPN essentially makes the public network as the user’s own private network instead of using dedicated connectivity. Virtual private networks (VPNs) provide the ability to create a secure network connection across a public network through the use of encryption and Firewall [40]. It’s necessary to note that the VPN itself has multiple implementations. VPN types include network-to-network, multiple service host-server, to single-service host-server. Each of these implementations can be used in a cloud computing environment, and each has security strengths and weaknesses [38][39].

4. PROPOSED METHODOLOGY

This research work deals with how to secure information using tables collide storage model to improve confidentiality when data is in rest.

4.1. Problem Formulation

Assume that there are N attributes in a table T1 under a database D. Split N attributes from table T1 into small pieces: N mod a small number (small number depends on total number of attributes in the database D) say P11,P12,P13,...,P1k1. Find P21,P22,P23,...,P2k2 from T2. Similarly, find pieces from T3...Tn.

<table>
<thead>
<tr>
<th>P11</th>
<th>P12</th>
<th>..........................</th>
<th>P1k1</th>
</tr>
</thead>
<tbody>
<tr>
<td>P21</td>
<td>P22</td>
<td>..........................</td>
<td>P2k2</td>
</tr>
<tr>
<td>..........................</td>
<td>..........................</td>
<td>..........................</td>
<td></td>
</tr>
<tr>
<td>Pn1</td>
<td>Pn2</td>
<td>..........................</td>
<td>Pnk1</td>
</tr>
</tbody>
</table>

Split() – split the tables into small pieces

D denotes Database to be stored in the cloud server. Ti denotes i_th table in D.

\[ n_i = \text{attribute-count}(T_i) \quad [1 <= i <= n_i] \]
r=random-number( ) [1<=r<=ni) ; piece = ni mod r
[1<=i<=ni] : if piece ==0 find another r

Kj=ni divided by piece; K1,K2...Kn are total number of pieces of T1,T2,..Tn in the database D.

Where attribute-count(Ti) function gives number of attributes in Ti AND random-number() function generates integer random number, Random number r will be generated till piece value is not equal to zero. When forming attributes of piece size, there will be Kj pieces in Ti, which is denoted in the figure 1.

Join() - collide pieces from different tables to achieve more confidentiality in the cloud storage.

Join pieces of different tables randomly to form collide small set of tables CT1, CT2,...CTm. Here m is the total number of small set of collide tables and m<=n.

CTk = Nk pieces of distinct( Pikj(Ti)) [1<=k<=m]

Store() - store m collide tables in distinct servers of public / private / hybrid service providers.

Retrieve () – retrieval of related pieces in correct order from cloud storage using private key table

C&C should maintain a table to identify all mixing as well as relationship among data in different servers at cloud storage. Here the table act as the private key. Using this key C&C retrieve required data for processing.

Table 2: Private Key Table Structure : No. of pieces in each table & Mapping of tables and servers

<table>
<thead>
<tr>
<th>Table id</th>
<th>No. of pieces</th>
<th>Mapping with table id</th>
<th>Cloud server id</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>K1</td>
<td>T1,P11,..,Tn,Pn1,..,CT1</td>
<td>Sn</td>
</tr>
<tr>
<td>T2</td>
<td>K2</td>
<td>T2,P21,..,Tn,Pn2,..,CT2</td>
<td>S1</td>
</tr>
<tr>
<td>T3</td>
<td>K3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tn</td>
<td>Kn</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The above table shows Private Key Table Structure.

5. RESULTS AND DISCUSSION

Table collide approach is compared with RAIN approach. Table collide model was implemented using Microsoft Azure, a public cloud service provider. Confidentiality of database increased with different usernames and passwords for each database. Collide tables were formed and placed in different regions. Private Key has formed which give information about how attributes were mapped. C&C would use this private key to retrieve data from different locations. Hackers cannot find entire information of a table on a single server and, moreover, the available information are a mixture of various tables and which are not related one. The information in cloud server are implicitly encrypted using table collide model. Table collide model ensures confidentiality by mixing and placing tables in different regions.
Figure 5. Servers Requirement

The chart shows that the accessing speed of Table collide approach is more than that of RAIN approach due to reduced number of servers on account of the mixing of data segment. The requirement of number of servers in existing approach is 300% more than that of the new approach. Accessing data from many servers causes delay in accessing process as where in Table collide accessing data needs 300% less number of servers when compared with the existing one. This increases the performance rate of accessing.

Figure 6. Utilization of Memory

Utilization of memory chart shows that Table collide approach uses the allotted memory 500% more than the existing approach. Table collide approach never leaves a large size of segment from a single table and it may have a large size of mixing of segments from various tables which are irrelevant for the hackers. Table collide approach effectively utilizes memory size allotted by servers and, at the same time, it ensures the confidentiality through mixing of irrelevant segments.

6. CONCLUSION AND FUTURE WORK

The result of implementation and comparisons prove that the storage model suggested here achieved more confidentiality through implicit encryption. Small number of pieces from different tables placed in regions with different usernames and passwords ensures secrecy. Frequently accessed tables can be kept in nearby regions to increase performance. Secret key table structure describes all collaborations and it can be maintained by C&C for processing. The result of chart ensures Table collide approach achieves both confidentiality and effective accessing speed. In future, time stamp can be calculated for accessing tables from different regions.

REFERENCES

6. B.Rajkumar, C.Yeo, S.Venugopal, S.Malpani, Cloud computing and emerging IT platforms: vision, hype, and reality for delivering
11. rivest, r.l., shamir, a., and adleman, l., "a method for obtaining digital signatures and public-key cryptosystems", Communications of the ACM, Vol 21, No. 2, February 1978, p. 120-26
34. McCabe M. Comment from the Conflicts of Interest, Privacy/Confidentiality, and Tissue Repositories: Protections, Policies, and Practical Strategies Conference co-sponsored by PRIM&R and the Columbia University Center for Bioethics. 2004 May 3-5; Boston, MA
52. Google (2011) Google Docs - Online documents, spreadsheets, presentations, surveys, file storage and more. Http://docs.google.com