International Journal of Computer Science
Oddity...Probe...Reviste...

http://www.ijcsjournal.com Volume 2, Issue 1, No 3, 2014. ISSN: 2348-6600
Reference ID: IJCS-041 PAGE NO: 229-235

Analysis of Security Issues and Privacy Preserving in Cloud Environment

S.Preetha, B.Sangeetha, Final Year M.C.A, Mrs. B. Usha Ranjini MCA., M.Phil,
Department of Computer Applications,V.S.B Engineering College,Karur, Tamil Nadu, India.
preethasmca@gmail.com,sangeevsb@gmail.com,usha.ren@gmail.com

***Abstract*— Cloud computing provides storage capacity without any infrastructure investment. Fair resource allocation among applications and dynamically adopts the load changes that means the resource allocation must be scalable both in number of machines and in number of sites. Preserving of privacy is still challenging problem in cloud. To overcome that, all the data are encrypted in existing system. But encrypting all data is neither efficient nor cost effective because it is very time consuming when the data are frequently accessed or processed. In this paper we propose one constraint based approach to identify which data need to be encrypted and which do not. It automatically reduces the time to access the data in cloud. Evaluation results demonstrate that the privacy-preserving cost of data can be reduced with our approach over existing ones where all data sets are encrypted.**

I. INTRODUCTION

Cloud consumers save their huge capital investment and concentrate on their core business. Therefore, many companies brought their business information into cloud. But many companies have not ready to import their business into cloud because of privacy concerns and security. The privacy concern caused by retaining data is most important but they are paid little attention. For that all the data are encrypted but it is very time consuming when data are accessed frequently. Each and every time the data is encrypt/decrypt. Thus, the cloud users select valuable information and store it to reduce the cost, by avoiding frequent re-computation to obtain these data .Data in cloud can be accessed by multiple parties and sometimes controlled by data owners. It somewhat reduces privacy issue. Most common approach to preserve the privacy is encryption.

Encrypting all data is adopted in current research but it is neither efficient nor time consuming when data are accessed frequently.

Most of the existing applications processed on unencrypted data only. Homomorphic encryption allows the process on encrypted data. But it is inefficient and expensive. Current privacy-preserving technique, generalization preserves most privacy attacks on one single data set. Preserving privacy for multiple data sets is not efficient. To preserve privacy of multiple data sets, first anonymize all data sets and decrypt them before storing them into cloud. Usually, data in cloud is very large. Hence, encrypting all data is inefficient when they are frequently accessed. We propose to encrypt part of data in cloud in order to reduce privacy preserving cost.

In this paper, we propose an approach to identify which data need to be encrypted and which do not. A tree structure is modeled from data sets to analyze privacy of data sets. Analyzing privacy leakage of multiple data sets is still challenging. For that, we design upper bound constraint to define the privacy. Based on such constraint, privacy preserving cost gets reduced. Finally, we design a heuristic algorithm to identify which data sets need to encrypted. Experimental results demonstrate that privacy preserving cost of data sets can be reduced with our approach over existing ones where all data sets are encrypted.

The major contributions of our approach

- We check the possibility of any privacy leakage without encrypting all data sets in cloud in order to reduce the privacy preserving cost.
- Heuristic algorithm is designed to identify which data sets need to be encrypted.
- Our approach reduces privacy preserving cost over existing approaches where all data are encrypted. I.e. we have to spend an amount on cloud for decrypting data.

II. RELATED WORKS

Data privacy in cloud is considered by most existing research. Encrypting data, provides security, but consumes more time and expensive. Puttaswamy et al. describes about a set of tools which identifies the data and encrypts from privacy. Ciriani et al. describes about an approach which

Published by SK Research Group of Companies (SKRGC).

International Journal of Computer Science
Oddity...Probe...Reviste...
http://www.ijcsjournal.com Volume 2, Issue 1, No 3, 2014. ISSN: 2348-6600
Reference ID: IJCS-041 PAGE NO: 229-235

combines encryption and data fragmentation. From this, privacy protection from the data storage is achieved. In order to achieve high security, the anonymization and encryption together have to be combined.

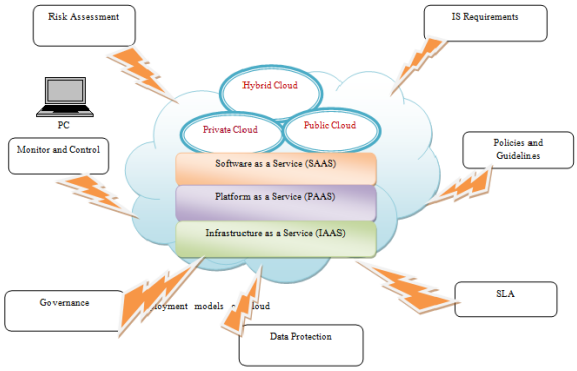
Wang et al. describes verifying data integrity with improved technique. Also a concept of Third Party Auditor (TPA) is utilized. From the client, no knowledge is gathered. TPA conducts auditing service, which is certified and trusted entity for conducting or requesting. The technique from TPA enables them to send the file and then it divides in to number of blocks.

Gartner identifies security issues to the cloud computing model in 2008. There are seven security issues that are to be addressed in a model. (1) User access information is transmitted through the internet with the high degree of risk-spends time and regulation as much as possible before assigning.(2) regulatory compliance from the third party organization-check security and clients. (3) Data location-some client might never know country or jurisdiction where the data are located. (4) Data segregation - encrypted information from multiple companies may be stored on the same hard disk. (5) Recovery- Every provider should have a disaster recovery protocol to protect user data. (6) Investigative Support – the suspects from the provider does not have any legal ways pursues for investigation.

(7) Long-term viability - refers to the ability to retract a contract and all data if the current provider is bought out by another firm

III. SECURITY ISSUES

Model, networking, platform, storage and software infrastructure are provided with ups and down which depends on the depicted figure below.



A. Private Cloud

In private networks, some vendors have recently used to describe the cloud computing. Internal enterprise data center is setup within an organization. The scalable resources and the virtual applications are pooled together by the cloud vendors. It is available for cloud user to share and use. It completely defers from public cloud, where the cloud resources and the application are managed by the organization which is similar to the intranet functionality. On the basis of utilization, private cloud is secured than private cloud due the internal exposure of private clouds. The private cloud must be operated only by organization and stake holders.

B. Public Cloud

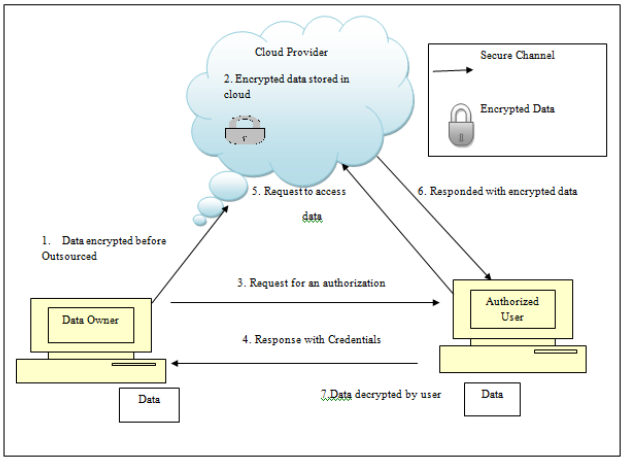
The traditional mainstream sense, whereby resources are dynamically provisioned on a fine-grained, self-service basis over the Internet, via web applications/web services, from an off-site third-party provider who shares resources must be described in a public cloud. It is based on the pay-per-use model similar to a prepaid electricity metering system which is flexible for spikes in demand for cloud optimization. It is less secure than the other cloud models, because it ensures all application and data access on the public cloud are not attacked by malicious.

C. Hybrid Cloud

It is similar to the private cloud which is linked to one or more external cloud services, centrally managed. Also described as a single unit and circumscribed by a secure network. It also provides the solution through the mix of public and private clouds. Hybrid provides more secure control of the data and allows various parties to access the information over the internet. The other management systems also interfere with the open architecture. It describes about the configuration by combining the local device with cloud services. It can also describe virtual and physical collocated assets. For example, the environment that needs physical servers or other hardware may act as a firewall or spam filter.

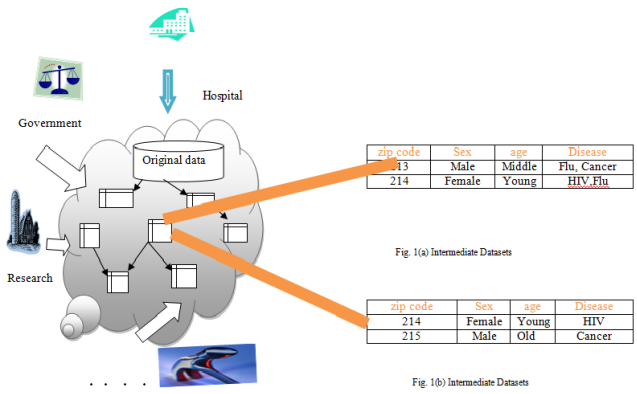
IV. SYSTEM DESIGN

International Journal of Computer Science
Oddity...Probe...Reviste...
http://www.ijcsjournal.com Volume 2, Issue 1, No 3, 2014. ISSN: 2348-6600
Reference ID: IJCS-041 PAGE NO: 229-235



User wants to access the data, first he/she checks whether the user is valid user or not. If the user is valid user, then he/she sends the request to access the data. He/she will be responded with the encrypted data which are requested. Finally, the data will be decrypted in user side.

V. SCENARIO



zip code	Sex	age	Disease
13	Male	Middle	Flu, Cancer
214	Female	Young	HIV,Flu

Fig. 1(a) Intermediate Datasets

zip code	Sex	age	Disease
214	Female	Young	HIV
215	Male	Old	Cancer

Fig. 1(b) Intermediate Datasets

Adversary (Knows that 25 years lady Affected by HIV)

One scenario is illustrated in the above figure. In the above figure, health service provider, e.g., Microsoft Health service provider moved their business information into cloud in order to reduce the cost of storage That means user can able to store the huge amount of data without any infrastructure investment. So that the consumer gets benefited. Microsoft Health service provider encrypts all original data sets for security. Data users like governments or research centre access the part of original data sets after anonymization. Intermediate datasets will be obtained during the data access. The intermediate datasets are often stored for data reuse and to reduce cost.

Here 1(a), 1(b) indicates two intermediate datasets. Fig. 1 satisfies 2-diversity, i.e., at least one field in each table must be same. Here, zip code=21400 is same in both intermediate data sets. It is known as quasi identifier. From that, adversary can able to conform that the lady is suffered from HIV who is living in the zip code 21400. To prevent that, any one of the intermediate data set will be encrypted. Assume the size of both intermediate data sets will be same but the frequency of accessing first intermediate data set will be 50 and the frequency of accessing the second intermediate data set will be 200. In that case, the first intermediate data set will be encrypted in order to confine privacy.

This situation is applicable, when the small number of intermediate data sets is generated. But in most of the real world applications, more number of intermediate data sets will be obtained. In that case, it is difficult identify which data sets need to be encrypted.

VI. PROBLEM DEFINITION

The problem resides in a shared environment, but the data owner should have full control over determining the right to access. They are allowed to gain access only once. It provides better data control in the cloud. It is based on the heterogeneous security approach from the essential element that also shifts the protection from system and applications. From this approach, the document must be self described and remain independent of environment. It also approach on cryptographic policy, then the policy rules must be considered. When someone authorizes the data, the system should check the policy to ensure the satisfaction of the rules. From this existing technique, we can utilize data security or privacy protection and significant computation attention. It addresses the literature from the data provenance issues. In such cases, information can be related to a particular hardware component which must be associated with a piece of data.

A. Problem Analysis

International Journal of Computer Science
Oddity...Probe...Reviste...
http://www.ijcsjournal.com Volume 2, Issue 1, No 3, 2014. ISSN: 2348-6600
Reference ID: IJCS-041 PAGE NO: 229-235

We defined several basic notations. Let d0 be a original data set and D denotes the set of intermediate data sets which are obtained from the original data set d0 i.e., D= {d1, d2,….., dn}.where n is the number of intermediate data sets.

1) Sensitive Intermediate data set Graph-SIG:

A DAG(Directed Acyclic Graph) representing the relationship between intermediate data sets D from d0 is defined as sensitive intermediate data set graph, specified as SIG. Here SIG = {V, E}, V = {d0} U D, E is a set of directed edges.

2) Sensitive Intermediate data set Tree-SIT:

The data in our research must be employed with the data provenance. Provenance can be derived from some objects and data which can be recognized by an organization. It can regenerate the data set from the data provenance. We assume that the information is recorded in the data provenance which is used to generate relationship. SIG defined as the sensitive intermediate data set tree structure represents the generation relationship and intermediate data sets. It also captures the propagation of privacy sensitive information. It is generally scattered to its offspring data sets. From this, we can analyze privacy disclosure of multiple data. We must present our approach to an SIT and then extend to SIG with minor modifications. It also satisfies certain privacy requirement which is assumed by an intermediate data. It results in violating the privacy requirement from the high risk of revealing the privacy sensitive information. Privacy leakage data can be obtained by the adversary form the privacy sensitive information.

The value of privacy leakage can be detected. It is challenging to acquire the exact value of privacy leakage due to the interference channel among the multiple data sets.

VII.HEURISTIC ALGORITHM

Description	Repeatedly identifies the intermediate data sets that need to be encrypted which reduces the privacy preserving cost
Input	A SIT with root d_0; all attribute values of each intermediate data set are given, i.e., size, frequency, privacy leakage, privacy requirement threshold(upper bound constraint)
Output	A vector of local solutions($\pi_{1,\ldots,}\pi_H$) that comprise a near optimal global privacy preserving solution; And the global privacy preserving cost

Step 1 Initialize the following variables.

1.1 Define a priority queue: PQueue.

1.2 Construct the initial search node with the root of the SIT; SN0= ((π0) <- ({d0},ϕ), f(SN0) <- 0, ED0 <- {d0}, Ccur <- 0, ϵ1 <- ϵ, i.e., the five parameters are the current solution , the current heuristic value, the current ED, the current cost and the privacy leakage requirement for the sequent layers.

1.3 Add the node into PQueue; PQueue <- SN0

Step 2 Repeatedly retrieve the search nodes from PQueue, and in turn add their child search nodes to PQueue

2.1 Retrieve the search node with the highest heuristics from PQueue: SN1 <- PQueue.

2.2 Check whether EDi = ϕ. If yes, a solution is found and the algorithm will go to step 3.

2.3 Label the datasets in CDEi as encrypted if their privacy leakage is larger than ϵi. Sort the unlabeled data sets in CDEi ascending according to Ck/PLs (dk), dk ϵ CDEi; SORT(CDEi). If the number of unlabeled datasets are larger than M, only the first M datasets are considered to generate candidate nodes.

2.4 Generate all the possible local solutions in Ai.

2.5 Select a solution from Ai; π <- SELECT (Ai).

1) Calculate the privacy leakage upper bound of this solution and the encryption cost:

PLlocal <- ∑d ϵUDπ PLs (d), Clocal <- ∑dkϵEDπ (Sk . CR. fk), where π = (EDπ, UDπ).

International Journal of Computer Science
Oddity...Probe...Reviste...
http://www.ijcsjournal.com Volume 2, Issue 1, No 3, 2014. ISSN: 2348-6600
Reference ID: IJCS-041 PAGE NO: 229-235

2) Calculate the remaining privacy leakage $\epsilon i + 1 <- \epsilon i - PLlocal$.

2.6 Compute the heuristic value

2.7 Construct new search node from the obtained values, add it to PQueue. Then go to Step 2.7.

Step 3 Obtain the global encryption cost Cglobal; Cglobal <- Ccur, and the corresponding solution ($\pi 1 \ldots \pi H$).

VIII. APPROACHES

A. Minimum Privacy Preserving Cost

The global encryption solution exists under the PLC constraints, because there are many alternative solutions to the problem. Each data set has various size and frequency of usage leading to overall cost with the different solutions. The compressed tree helps to categorize the data in the generated type value. This is calculated only for the layer level value less than the threshold value. It must be omitted when the field value is greater than the threshold value. There are some restrictions allowed to identify the privacy preserving cost value. The minimum value from the data sets under privacy leakage is dais. Size, price, allocated transaction in GB or MB will be identified by this value. It finds the record under the field where the values are iterative. There is no further elimination in the module from the data set size. Then we can identify the minimum privacy preserving cost. Pseudo is the minimum solution as upper bound of joint privacy leakage is approximated to exact value. It is necessary to use heuristic algorithm for scenario where we can obtain a large number of intermediate data sets. It is also used to obtain an optimal solution with a high efficiency.

B. Anonymization and Encryption

The process of converting the data set is used to store at cloud storage space. The analysis and search of encrypted strings with an integer value is waste of time. From that we cannot categorize the integer values. So, we can solve the problem by using technique anonymization. The privacy preserving cost must be reduced by both encryption and anonymization for the data set. The data which is encrypted and anonymzed stored in a cloud space that can be easily transferable. Then, finally compare the results with the existing graph. It also produces the graph for heuristic privacy preserving cost value. User login is used to view the data set by the data holder that produces the anonymized and encrypted data sets. If the number of intermediate data set is larger, then we call both CALL and CHEU. For the large number of intermediate data sets cost will be effective. CALL is proportional to the number of intermediate data sets which increases notably. CHEU also increases with the number of data sets that are required to encrypt. CHEU also drops when the privacy leakage becomes larger then CALL keeps invariable.

.

IX. CLOUD COMPUTING CHALLENGES

There are numerous challenges associated with the current adoption of cloud computing because users are skeptical about their authenticity. The prevention of cloud computing from being adopted by an organization was based on the survey conducted by IDC in 2008. The challenges are described below:

A. Security

Security issue plays an important role in hindering cloud computing acceptance. By following this technique, putting your data, running your software on someone else's hard disk using someone else's CPU appears to many. In the organizations, data and software has several security issues, such as phishing, botnet (running remotely on a collection of machines). The new security challenges are multi tenancy model and pooled computing resources in cloud computing. It requires the novel technique to tackle with it. For example, hackers can use Cloud to organize botnet as Cloud often provides more reliable infrastructure services at a relatively cheaper price for them to start the attack.

B. Costing Model

The communications, integration, tradeoffs must be consider to cloud consumers. It can reduce the infrastructure cost while migrate to the cloud, then it does not raise the rate of data communication. The data from public and community cloud from the computing resources are likely to be higher. The main problem particularly depends upon the consumer who uses the hybrid cloud deployment model. The organization data is distributed to public or private community cloud based on demand computing which makes sense to the CPU intensive jobs.

C. Charging Model

The elastic resource is more complicated than the regular data centers. It also calculates cost based consumption of static computing. For SaaS cloud providers,

Published by SK Research Group of Companies (SKRGC).

International Journal of Computer Science
Oddity...Probe...Reviste...
http://www.ijcsjournal.com Volume 2, Issue 1, No 3, 2014. ISSN: 2348-6600
Reference ID: IJCS-041 PAGE NO: 229-235

the deployment of multi-tenancy within their offering of cost should be substantial. It includes re-design and redevelopment that can be used by single-tenancy model. The new feature which provides intensive customization, performance and security enhancement for user access are induced by the above changes. The weigh up tradeoff between multi-tenancy provision and cost savings can be yielded through overhead amortization, reducing a number of license, etc., In SaaS cloud providers, it also provides profitability and sustainability in a charging model.

D. Service Level Agreement (SLA)

Computing resources do not have control over the cloud customer. It needs to ensure the quality, availability, reliability and performance. It consumes core business function onto their entrusted cloud.

It provides service delivery to obtain the guarantees from the consumer. Service legal agreements can be negotiated between the providers and consumers. SLA can be specified with level of granularity, tradeoff between expensiveness and complicate, so they can cover most of the consumer expectation. It is relatively simple to be weighted, verified, evaluated and enforced by a mechanism in the cloud. In addition to that, different meta-specification should be defined with different client offerings (IaaS, PaaS, and SaaS). It also increases the problem for cloud providers to implement. It needs to develop the mechanism for user feedback and customization features in to the SLA evaluation framework.

E. What to migrate

Based on a survey (Sample size = 244) conducted by IDC in 2008, the seven IT systems/applications being migrated to the cloud are:

IT Management Applications -26.2%,
Collaborative Applications -25.4%,
Personal Applications -25%,
Business Applications -23.4%,
Applications Development and Deployment-16.8%,
Server Capacity -15.6%, and
Storage Capacity -15.5%.

From this result, security or privacy concern in moving their data on the cloud should be organized. There are some IT systems such as peripheral functions which are easy to move. When an employing IaaS which is compared to SaaS are being conservative. Outsourced cloud and core activities are the marginal functions which are kept in house. From the survey, nearly 31% of the organization moves the storage capacity to the cloud. There are more collaborative application compared with old applications.

F. Cloud Interoperability Issues

Based on a survey (Sample size = 244) conducted by IDC in 2008, the seven
IT systems/applications being migrated to the cloud are:

IT Management Applications -26.2%,
Collaborative Applications -25.4%,
Personal Applications -25%,
Business Applications -23.4%,
Applications Development and Deployment-16.8%,
Server Capacity -15.6%, and
Storage Capacity -15.5%.

From this result, security or privacy concern in moving their data on the cloud should be organized. There are some IT systems such as peripheral functions which are easy to move.

When an employing IaaS which is compared to SaaS are being conservative. Outsourced cloud and core activities are the marginal functions which are kept in house. From the survey, nearly 31% of the organization moves the storage capacity to the cloud. There is more collaborative application compared with old applications.

X. CONCLUSION

In this paper we address the security issues and proposed an approach that identifies which intermediate data set need to be encrypted and which other does not. A sensitive intermediate data set tree is designed to define relationship between intermediate data sets. We have designed the practical heuristic algorithm. Heuristic algorithm is designed in order to identify which data sets need to be encrypted. Evaluation results have demonstrated the cost of preserving privacy can be significantly reduced over existing ones where all intermediate data sets are encrypted.

REFERENCES

[1] M. Armbrust, A. Fox, R. Griffith, A.D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, “A View of Cloud Computing,” Comm. ACM, vol. 53, no. 4, pp. 50-58, 2010.

International Journal of Computer Science
Oddity...Probe...Reviste...
http://www.ijcsjournal.com Volume 2, Issue 1, No 3, 2014. ISSN: 2348-6600
Reference ID: IJCS-041 PAGE NO: 229-235

[2] R. Buyya, C.S. Yeo, S. Venugopal, J. Broberg, and I. Brandic, "Cloud Computing and Emerging It Platforms: Vision, Hype, and Reality for Delivering Computing as the Fifth Utility," Future Generation Computer Systems, vol. 25, no. 6, pp. 599-616, 2009.

[3] L. Wang, J. Zhan, W. Shi, and Y. Liang, "In Cloud, Can Scientific Communities Benefit from the Economies of Scale?," IEEE Trans. Parallel and Distributed Systems, vol. 23, no. 2, pp. 296-303, Feb. 2012.

[4] H. Takabi, J.B.D.. joshi, and G. Ahn, "Security and Privacy Challenges in Cloud Computing Environments," IEEE Security & privacy, vol. 8, no. 6, pp. 24-31, Nov. /Dec.
2010.