PRIVACY PROTECTION OF FREQUENTLY USED DATA SENSITIVE IN CLOUD SEVER

T.Pavithra
M.E Computer Science And Engineering
S.A.Engineering College,
Chennai-600077, Tamil Nadu.

Mrs.G.Umarani Ph.D
Professor & Head
S.A.Engineering College,
Chennai-600077,Tamil Nadu,

***Abstract---The cloud computing is one of* the most innovative technology for to storing the information to the cloud server. So, now a days many organization moving to utilizing the cloud for economical benefits. The data owners must place their valuable sensitive data's into the public cloud servers which are not within their trusted domains. Privacy-preserving database in cloud allows a database owner to outsource its encrypted database to a cloud server. Hence, security and privacy of data is the major concern in the cloud computing. To overcome this information disclosure, the service providers encrypts the sensitive data before uploading to the cloud servers. The plain text keyword search is impossible, because data stored in public clouds increases in an exponential manner. Previously encrypt the data is based on the frequency of queries on the table. The data which are common among the tables can be retained and the sensitive information can be encrypted. By using the partial inference control based data disclosure approach the cost and delay for computation can be reduced. Then we can apply the RSA algorithm for encryption, is used to achieving the high privacy . In addition to predicting the data cannot be done at any case. This provides a high level security at low cost .**

***Index Terms*-Inference control, RSA algorithm Cloud computing, Time flow scheduling, Encryption.**

I. INTRODUCTION

Cloud Computing is the fastest growing technology which is provide the storage and resource sharing options for the users. Nowadays every technology moving to the cloud. Today medical information's also moving to the cloud. Moving the sensitive information's to the cloud, having the lots of security issues. Personal health information system is the system for ensuring privacy preserving the system and distribution of medical information to the authorized users Cloud Computing is a distributed service through which we can easily sharing the resources or information, software on your computer or other devices via the Internet connection. This means that we can access the information we want anytime, anywhere, so we can achieve less time to accessing the dataset from the cloud. Cloud Providers offer services that can be grouped into folloing categories.

A. Cloud Computing Models

The "cloud" in cloud computing can be defined as the set of hardware, networks, storage, services, and interfaces that combine to deliver aspects of computing as a service. Cloud services include the delivery of software, infrastructure, and storage over the Internet based on user demand.

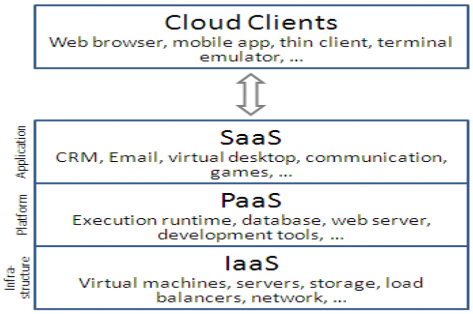
Cloud Clients
Web browser, mobile app, thin client, terminal emulator, ...
Application
SaaS
CRM, Email, virtual desktop, communication, games, ...
Platform
PaaS
Execution runtime, database, web server, development tools, ...
Infra-structure
IaaS
Virtual machines, servers, storage, load balancers, network, ...

Fig. 1. Cloud Computing Models

Software as a Service (SaaS)

Software as a Service (SaaS) is a software distribution model in which applications are hosted by a vendor or service provider and made available to customers over a network, typically the Internet

Platform as a Service (Paas)

Platform as a Service (PaaS) is a way to rent hardware, operating systems, storage and network capacity over the Internet. The service delivery model allows the customer to rent virtualized servers and associated services for running existing applications or developing and testing new ones.

Infrastructure as a Service (Iaas)

Infrastructure as a Service(IaaS) is a provision model in which an organization outsources the equipment used to support operations, including storage, hardware, servers and networking components. The service provider owns the equipment and is responsible for housing, running and maintaining it. The client typically pays on a per-use basis

B. Types Of Clouds

With cloud computing technology, large pools of resources can be connected through private or public networks. This technology simplifies infrastructure planning and provides dynamically scalable infrastructure for cloud based applications, data, and file storage. Businesses can choose to deploy applications on Public, Private, Hybrid clouds or the newer Community Cloud.

Public Cloud in public cloud storage, it can access by any subscriber with an internet connection and access to the cloud space.

Private Cloud

In private cloud storage, it is established for a specific organizations and limits to access to those organizations.

Hybrid Cloud

In hybrid cloud storage, it is combination of the public and private cloud storage. It means where critical cloud data located in private cloud while other data is stored and accessed from public cloud.

Community Cloud

In community cloud is a is a multi-tenant cloud service model that is shared among several or organizations and that is governed, managed and secured commonly by all the participating organizations or a third party managed service provider.

The owner of the data after uploading their information into the cloud, they accessing that data, the intermediate data set will be created, the privacy of the data is major concern, so to overcome this problem. we going to apply encryption techniques for maintaining the privacy. Previously they using some algorithm for encryption, but now I going to apply the RSA algorithm for encryption. This algorithm is simple and more effective to comparing other algorithms like message digest5 algorithm

II. RELATED WORKS

To clearly analysis the explore on privacy protection in cloud, for intermediate dataset to furnish the privacy preserving for sensitive data. Now, encryption is exploited by most existing re-search to ensure the data privacy in cloud [1]. To identify the security of the cloud, how they will provide the security to our data's and finding the uniqueness of the security, then we come to know how to eliminating the potential threats for getting the better solution this information's are referred in [2]. The intermediate dataset will generated for after searching and accessing the original data, then these data's are used to retrieve the deleted data, it will act as a backup referred in [3]. To generate the intermediate data dependency graph(IDG), for regenerate the deleted intermediate dataset items, and frame the novel algorithm for to find the minimum cost storage for dataset items. Although encryption is well for data protection in these approaches, it is very vital function to encrypt and decrypt often in data- sets for many applications. Encryption is usually integrated with other function to achieve cost minimization, high data utilization and privacy protection. [4] To frame a system named like *Sedic* which is used to partitioning the security label of sensitive data in Mapreduce job computing and then uploading the data to public cloud, which does not contain the sensitive data. The sensitivity of data is very vital to make sure that the above labeled approaches available [5]. proposing an approach that data fragmentation and integrate encryption for to get protection of sensitive data for distributed data storage and encryption of the datasets is one part. We update this line, but integrate encryption and data anonymization together to fulfill cost-effective privacy security. The Vital function of retaining intermediate datasets in cloud has been widely recognized [6], but the issues of processing on privacy, incurred by such datasets just commences. Davidson et al. [7], [8], studied the problem in workflow derivation, and apply to achieve high value of provenance information and privacy preserving of module via whipping carefully a subset of inter-mediate data.

This general method is similar to us, yet our research mainly concern on data security preserving of sensitive data from an cost economical perspective while theirs mainly concentrates on functionality privacy of load workflow dataset modules, but does not give that much importance to data privacy. their research also differs from ours in many aspects such as privacy quantification , cost models and data hiding techniques. Their research is mainly concentrate on retrieving the hided data, but our research is to overcome the cost of the anonymized data.

The PPDS (privacy preserving data set) research has mainly investigate extensively on privacy-preserving data set issues and make fruitful improvement with a stream of reserving methods and privacy models in reference paper [9]. Privacy function such as *l*- diversity [11] and *k*-anonymization [10] are both function to model and quantify privacy, but most of the people apply the single data set item. We also proposing a Privacy principles for multiple datasets, but they aim to achieve a specific scenarios such as sequential data releasing or continuously publishing a data. The analysis of paper [12], [13] feat information theory to measure the privacy via to achieving the maximum encryption principle [14]. The privacy quantification this process is referred in [12], [13]. More encryption techniques are like a simplification has been proposed for to preserving the privacy, but these methods are only fail to overcome the difficulty of preserving privacy for multiple sensitive datasets items. This approach include encryption with anonymization and overcome those problems for to achieve privacy preserving of multiple sensitive datasets items.

III. EXISTING SYSTEM

In the existing system, we study the strategies for efficiently achieving data staging and data storage for privacy concern on a set of vantage to reduce the computational cost of encryption or decryption of data sets in a cloud system with a minimum outlay. Surplus data used for improvising the efficient optimal solutions is based on the dynamic upper bound privacy which is polynomial bounded by the number of service requests and the number of distinct data items in cloud. In existing system they used the following concepts.

A. Novel upper bound privacy leakage constraint- based approach

This approach is used to select the highly valuable subset of intermediate datasets that are very important to encryption for minimizing privacy-preserving cost. The main function is to finding the vital data for encryption by using some calculation.

B. Privacy-Preserving Cost Reducing Heuristic Algorithm

To design a heuristic algorithm for reduce privacy-preserving cost. In the state-search space for a *SID (*sensitive intermediate dataset*)*, a state node the layerhere in refers to a vector of partial local solutions, but the height is the same reducing the cost for encrypted data's. It's to find the near optimal solution in a limited search space.

This is partial as most of the existing staging or privacy upper bound targets towards a class of services that access and process the decrypted data and thereby inherit the severity of data when access time sequence is more. Alternatively, a constraint optimization problem can be defined as a regular constraint to find a solution to the problem whose cost, evaluated as the sum of the cost functions, is minimized. Third parties who have privilege over intermediate datasets are created in order to reduce the frequent access of data from cloud directly that increases the cost. Hence the procedure of anonymization andhomomorphic type of encryption are done in the system. In turn, avoids the possibility of inference channel analysis.

IV. PROPOSED SYSTEM

This proposed system is designed to identify only the important and critical intermediate datasets that needs to be encrypted for security purposes, hence reducing encryption/decryption cost and thus maintaining data privacy. One way for evaluating this upper bound for a partial solution in our existing paradigm is to consider each constraint separately and mining the data in order to restrict access when the user claims to find the original information. For each constraint, the maximal possible value for any of these values is an upper bound may recover privacy-sensitive partial column level encryption. Hence an column wise encryption in the unencrypted data's of intermediate datasets are proposed.

A. Representative pattern frequent mining algorithm

This algorithm is used to Finding the frequently used vital data for the encryption based on the counting of access in particular dataset. Then also this algorithm to identifying the sensitive dataset from the original data set. The purpose of to identifying sensitive data set is to reducing the encryption cost in the cloud. Because, we going to encrypt this data particular data only.

B. Time scheduling algorithm

To frame the Flow time scheduling algorithm for data. Flow time scheduling algorithm is mainly used to maintain a log based tracking for frequent and un frequent usage of data under the time criteria. This algorithm automatically to schedule the newly updated data for same process and finding the frequent data and also checking the, if data is already anonymized or not.

C. RSA algorithm

I. Choose two prime numbers p and q.
II. Find n such that n = pq.n will be used as an both public and private keys.
III. Find the totient of n, $\phi(n)$.
$\phi(n)=(p-1)(q-1)$.
IV. Choose an e such that $1 < e < \phi(n)$, and such that e and $\phi(n)$ share no divisors other than 1(e and $\phi(n)$ are relatively prime).e is kept as the public key .
V. Determine d (using modular arithmetic) which satisfies the congruence relation.
$de \equiv 1 \pmod{\phi(n)}$.
In other words, pick d such that de - 1 can be evenly divided by (p-1)(q-1), the totient, or $\phi(n)$.
since $\phi(n)$ and e are relatively prime and d is to be the modular multiplicative inverse of e.
d is kept as the private key .
The public key has modulus n and the public (or encryption) key e. The private key has modulus n and the private (or decryption) key d, which is kept secret.

V. SYSTEM ARCHITECTURE

System architecture is the conceptual model that defines the structure and/or behavior of the system. The system architecture for the proposed system is given in the Figure2. It provides a way in which data will be selected for encryption and stored in cloud. The architecture diagram shows that Data owners to uploading their valuable information to the cloud server. The user can searching and accessing the desired information, now to using the inference control concept is able to finding the most searchable data Then decide which data going to be encrypted and which no need is. The unauthorized user decide to accessing those data, if the data is not encrypted the user able to accessing the data, but the data is encrypted the user cannot accessing the data, the access will be denied. This architecture diagram clearly explain, how the data will be stored in the cloud and how encrypting the sensitive data for maintaining the privacy.

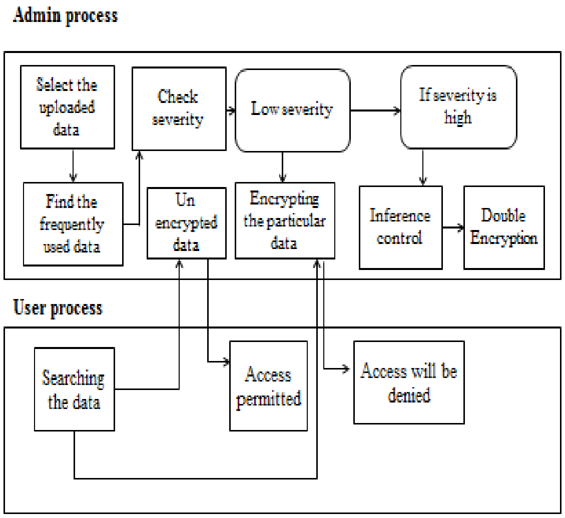

Fig. 2. System Architecture

A. Frequent pattern Data Sets Identification

The parameter that we have to decide upon is called support of an item set. In order to identify the item sets that are accessed most commonly in the cloud storage environment. Pattern is a form or model (or, more abstractly, a set of rules) that can be used to make or to generate things or parts of a thing. These frequently accessible datasets are stored in the intermediate datasets of cloud storage so as to avoid the cost of accessing the cloud frequently.

B. Relational Data Sets Identification

The entity is a person, object, place or event for which data is collected. The relationship is the interaction between the entities. The table with common references with another tables are identified so as to prevent the privacy leakage through inference channel analysis. Inference channel analysis is a control used in the output of databases to stop a person who has access to only summary information from being able to determine (infer) a particular value for a particular record. The severity estimation of the datasets are identified in order to perform inference channel analysis to avoid privacy leakage.

C. Encryption of Data sets

Anonymity typically refer to the state of an individual's personal identity, or personally identifiable information, being publicly unknown. This process of anonymization is done for maintaining privacy as well as to reduce the utilization size of the data. As the size of the data utilization reduces cost of the utilized data get reduced. Here we just encrypting the selected particular data ,the anonymization is also referred to as encryption. Now the encrypted data are utilizing less space from the cloud , so the cost of to maintaining the privacy of the data will get reduced.

D. Double Encryption

The use of encryption/decryption is as old as the art of communication. In wartime, a cipher, often incorrectly called a code. Encryption is the conversion of data into a form, called a ciphertext, that cannot be easily understood by unauthorized people. The anonymized datasets in the previous module are checked for its relativity with other datasets belongs to another

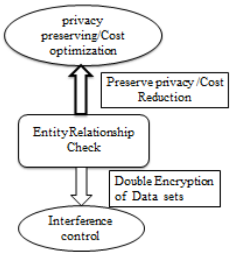
privacy preserving/Cost optimization
Preserve privacy /Cost Reduction
Entity Relationship Check
Double Encryption of Data sets
Interference control

Fig 3. Double Encryption

VI. CONCULUSION

This paper, to proposed an approach that mainly used to furnish the privacy of data. Then overcome the problem for storing the data in to cloud server and how to secure our data. This paper, mainly to considering the cost of encrypted data, that are stored in the cloud server. The RSA algorithm used to annonymizing the selected sensitive data only, so we limited amount and secure highly vital data.

REFERENCES

[1] H. Lin and W. Tzeng, "A Secure Erasure Code- Based Cloud Storage System with Secure Data Forwarding," *IEEE Trans. Parallel Distrib. Syst.*, vol. 23, no. 6, pp. 995-1003, 2012.

[2] D. Lekkas and D. Zissis, "Addressing Cloud Computing Security Issues," *Fut. Gener. Comput. Syst.*, vol. 28, no. 3, pp. 583-592, 2011.

[3] S.Y. Ko, I. Hoque, B. Cho and I. Gupta, "Making Cloud Intermediate Data Fault-Tolerant," *Proc. 1st ACM Symp. Cloud Computing (SoCC'10)*, pp.181-192, 2010.

[4] Y. Yang, D. Yuan, , X. Liu and J. Chen, "On- Demand Minimum Cost Benchmarking for intermediate Dataset storage in scientific cloud workflow system," J Parallel Distributed computing, vol. 71, no. 2, pp. 316-332, 2011

[5] K. Zhang, X. Zhou, Y. Chen, X. Wang and Y. Ruan "Sedic: Privacy-Aware Data Intensive Computing on Hybrid Clouds," *Proc. 18th ACM Conf. Computer and Communications Security (CCS'11)*, pp. 515-526, 2011

[6] V. Ciriani, S.D.C.D. Vimercati, S. Foresti, S. Jajodia, S. Paraboschi and P. Samarati, "Combining Fragmentation and Encryption to Protect Privacy in Data Storage," *ACM Trans. Information and System Security*, vol. 13, no.3, pp. 1-33, 2010.

[9] B.C.M. Fung, K. Wang, R. Chen and P.S. Yu, "Privacy-Preserving Data Publishing: A Survey of Recent Developments," *ACM Comput. Surv.*, vol.42, no. 4, pp. 1-53, 2010.

[10] P. Samarati, "Protecting Respondents' Identities in Microdata Release," *IEEE Trans. Knowl. Data Eng.*, vol. 13, no. 6, pp. 1010-1027. 2001.

[11] A. Machanavajjhala, D. Kifer, J. Gehrke and M. Venkitasubramaniam, "*L*-Diversity: Privacy Beyond *K*-Anonymity," *ACM Trans. Knowl. Discov. Data.*, vol. 1, no. 1, Article 3, 2007.

[12] G. Wang, Z. Zutao, D. Wenliang and T. Zhouxuan, "Inference Analysis in Privacy-Preserving Data Re-Publishing," *Proc. 8th IEEE Int'l Conf. Data Mining (ICDM '08)*, pp. 1079-1084, 2008.

[13] W. Du, Z. Teng and Z. Zhu, "Privacy-Maxent: Integrating Background Knowledge in Privacy Quantification," *Proc. ACM SIGMOD Int'l Conf. Management of Data (SIGMOD'08)*, pp. 459-472, 2008.

T.Pavithra currently studying PG in Computer Science at S.A.Engineering College, Chennai. She completed her UG in Phonnaiyah Ramajam College of Engineering and Technology , Thanjore. My areas of interest are Network Security, Data Structures and Cloud Computing.

Dr.G.Umarani Srikanth is currently working as a Professor and Head, Department of PG Studies, S.A.Engineering College, Chennai, India. She has 17 years of teaching experience. Her areas of interests include Compilers, Theory of Computation, Data Structures and Soft Computing.