International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249 – 8958 (Online), Volume-9 Issue-4, April, 2020

Cloud Based Privacy Preserving Dynamic Intermediate Datasets

V.Sarala, P.Shanmugapriya

Abstract: ***Design an encryption of privacy preserving and scheduling of intermediate datasets in cloud. Implemenation of encryption is done as follows: Identification of intermediate datasets that needs to be encrypted. Based on frequent pattern mining the least frequent intermediate datasets are encrypted. Perform column level encryption to the sensitive information. Predicting the data based on inference analysis would not be possible. So that the data will be secure when compared to the existing system.***

Keywords–Automatic scheduling, Intermediate datasets, Privacy preserving

I. INTRODUCTION

To reduce or to save the cost of privacy-preserving of intermediate datasets in cloud, from the original datasets or the public cloud the intermediate datasets are generated through the proper authorization, with that intermediate datasets third party can access modify the data analyse, reanalyse the result and save the modified data .Encrypting ALL datasets is time consuming. Based on privacy leakage upper bound based constraint and the representative frequent pattern matching algorithm ,find the least and most frequent table from the datasets. The Least frequent intermediate datasets is anonymised and encrypted using homomorphic encryption. All these are analyzed by privacy leakage upper bound based constraint .Encrypting a part of intermediate datasets will save the cost of re-computing and privacy is maintained. When this encryption is not done Inference analysis would be possible therefore the sensitive information can be easily predicted by comparing multiple intermediate datasets or from the public datasets and it is considered to be static. Therefore encrypting a part of intermediate datasets would save the privacy preserving cost of the data in cloud.

II LITERATURE REVIEW

Sensitive information is kept confidential by encrypting datasets. Inference analysis is made from unencrypted datasets. Prediction of data would be possible. Size and frequency of data is not dynamic in nature [1]. Decentralized erasure code method is used for Datasets. Storage server and key server are two systems used for storage system. Data is more confidential and highly protected security.

Encryption is done to all data stored in cloud and also high computation cost [2]. Problem searches over encrypted information. Keyword search over encrypted data by multiple users and search is based on attributes [3].Minimum cost strategy for Intermediate datasets, Low computation cost and storage cost is the advantage . Statistic intermediate dataset is available [4].Sensitive data stay on private cloud. We can find the minimum cost Intermediate datasets storage strategy. Users split their computing tasks. Bulk of data can be split into small data [5].

Revised Manuscript Received on April 25, 2020.
* Correspondence Author

V.Sarala*, Department of computer science,Meenakshi college of Engineering,Chennai, IndiaSaralapurush@gmail.com

Dr.P.Shanmugapriya, Department of Computer science,SCSVMV University,Kanchipuram,IndiaPriya_prakasam@yahoo.co.in

III EXISTING SYSTEM

System encrypting a part of intermediate datasets while retaining the other datasets based on upper bound privacy leakage constraint approach, Encrypting ALL datasets is time consuming .Size and Frequency of data is also static.To identity the leakage Heuristic algorithm is used in Existing system.Figure 1 represent the Block diagram of Existing system.

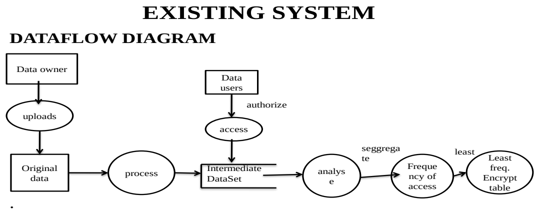


Figure 1

IV PROPOSED SYSTEM

Proposed system is being used for the dynamic datasets in cloud. This provides a highest level of security to the system. While accessing the table already encrypted data will become the high frequent access table and the table with the high frequent will become least access table .At that time dynamic scheduling is used to monitor the usage of data.Figure 2 represent the Block diagram of proposed system.

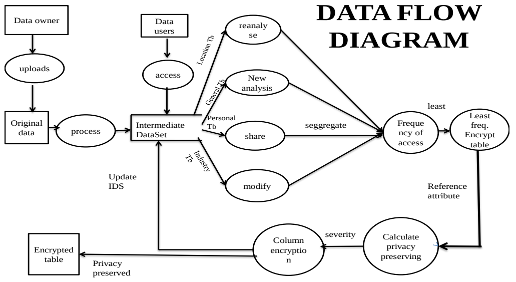


Figure 2

Retrieval Number: D8380049420/2020©BEIESP
DOI: 10.35940/ijeat.D8380.049420
Journal Website: www.ijeat.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication
© Copyright: All rights reserved.

V MODULE DESCRIPTION

Module description contains four modules. They are

- Creation of datasets and Least frequent pattern Data Sets Identification
- Entity Relational Data Sets Identification and encryption(AES)
- Scheduling of intermediate datasets
- MD5 with Triple DES

5.1.1 Representative Pattern Frequent Mining Algorithm:

In Creation of datasets and Least frequent pattern Data Sets Identification, original datasets for a Government application where all the people related information is present are uploaded into the cloud by the data owner. When processing these original datasets the data owner store the valuable intermediate data sets from that the least frequent table is encrypted by advanced encryption standard algorithm. Figure 3 represent the example of Frequent pattern mining algorithm.

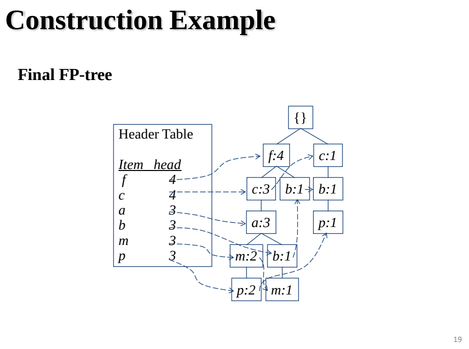

Construction Example

Figure 3

5.1.2 Advanced Encryption Standard Algorithm (AES):

AES is a non-Feistel cipher that encrypts and decrypts a data block of 128 bits. It uses 10, 12, or 14 rounds. The key size, which can be 128, 192, or 256 bits, depends on the number of rounds.

5.1.3 Intermediate Datasets Scheduling Algorithm

The intermediate Datasets Scheduling Algorithm is implemented to schedule the intermediate datasets based on the frequency of accessing the datasets. This algorithm automatically updates each time whenever the datasets are accessed by the data users. This algorithm identifies the least and the most frequent intermediate datasets to perform the encryption process.

Algorithm for scheduling Algorithm

```
Declare Type=Dynamic
 **intialize** count=0,
 Access the intermediate Datasetss
**Update** count for a accessing the particular  file in the database
 **Begin**
Arrange the higher values to most frequent table
Arrange the least access count datasetss to least frequent table.
**End**
if size of table is added
**Begin**
 Add to the new user registration
 **End**
  **Else**
  **Begin**
Proceed the scheduling process based on access count.
  **End**
```

5.1.4 MD5 With Triple Des Algorithm

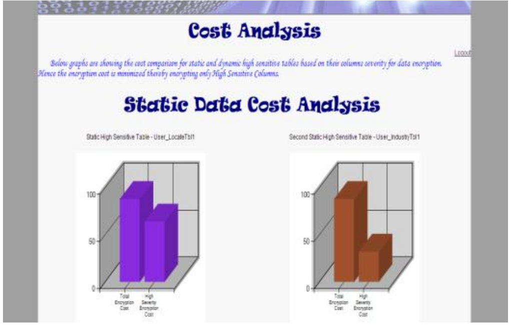
MD5 message-digest algorithm encrypt all passwords in database using some standard cipher,,it is also one way hashing algorithm.

Triple DES Algorithm:

In cryptography, **Triple DES(** Triple Data Encryption standard) algorithm block cipher, it consist of 3 steps first encrypting, and decrypting and again encrypting using the keys k1,k2 and k3 for encryption, whereas for decryption, first decrypting and encrypting and again decrypting using the 3 keys.

VI EXPERIMENTAL RESULTS

Privacy preserving cost is reduced when compared to encrypt all the intermediate dataset. To verify the performance, System is compared with the computational cost for encryption/decryption process, by encrypting a part of an intermediate datasets with all the intermediate datasets that are encrypted. Figure 4 and Figure 5 represent static and dynamic data cost analysis.

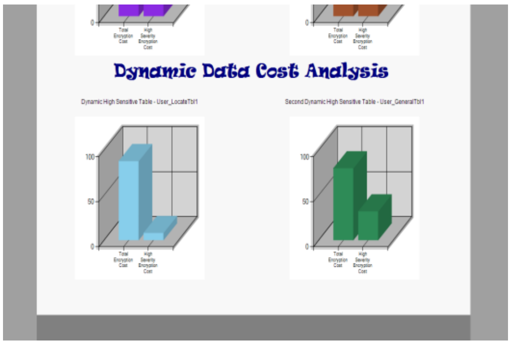

Cost Analysis

Below graphs are showing the cost comparison for static and dynamic high sensitive tables based on their columns severity for data encryption. Hence the encryption cost is minimized thereby encrypting only High Sensitive Columns.

Static Data Cost Analysis

Logout
Static High Sensitive Table - User_LocateTbl1
Second Static High Sensitive Table - User_IndustryTbl1
100
50
0
Total Encryption Cost
High Severity Encryption Cost

Figure 4


Dynamic Data Cost Analysis

Total Encryption Cost
High Severity Encryption Cost
Dynamic High Sensitive Table - User_LocateTbl1
Second Dynamic High Sensitive Table - User_GeneralTbl1
100
50
0

Figure 5

Retrieval Number: D8380049420/2020©BEIESP
DOI: 10.35940/ijeat.D8380.049420
Journal Website: www.ijeat.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

VII CONCLUSION

The secure encryption of privacy preserving intermediate dataset enables to encrypt a part of intermediate datasets rather than encrypting all the datasets. This identifies the privacy sensitive information and therefore column level encryption is done to the intermediate dataset to prevent the revealing of privacy sensitive information. Automatic Dynamic Scheduling of intermediate dataset used to monitor the usage of data. Therefore the privacy preserving cost for the encryption/decryption process is reduced.

REFERENCES

1 A Privacy preserving upper bound constraint by Xuyun Zhang, Chang Liu, Surya Nepal, Suraj Pandey, IEEE transactions on distributed systems, vol. 24, no. 6, june 2013

2 A Secure Erasure code based cloud storage by H. Lin and W. Tzeng IEEE Trans. Distributed Systems, vol. 23, no. 6, pp. 995-1003, June 2012.

3 "Authorized Private Keyword Search Computing," by M. Li, S. Yu, N. Cao, and W. Lou, Proc. 31st Distributed Computing Systems 2011

4 On Demand minimum cost for Intermediate dataset by D. Yuan, Y. Yang, X. Liu, and J. Chen J. Parallel Computing, no. 2, pp. 316-332, 2011

5 Sedic:Privacy-Aware Data Intensive Computing by K. Zhang, X. Wang X. Zhou, , and Y. Ruan, 18th ACM Conf. Computer and Comm. Security (CCS '11),pp. 515-526, 2011.

AUTHORS PROFILE

V.Sarala ,has completed M.E in computer science and pursuing PhD in computer Engineering from SCSVMV university Kanchipuram.She has total 13 years of experience in teaching and working as a Asst.prof at Meenakshi college of Engineering Chennai.

Dr.P.Shanmugapriya,M.E.Ph.D in computer Science Engineering and working as a Associate professor in Computer science Department for SCSVMV university.She has 16 years of experience in teaching.