Price-effective Preserving Privacy of Intermediate Data Sets and Verifying Data in Cloud

[1] Amruta S. Patil, *[2] Prof.S.P.Patil, *[3] Prof. Suhel Sayyad
[1]ME (COMP) Appearing, [2]PHD (COMP) Appearing, [3]PHD (COMP) Appearing
[1]Department of Computer Science & Engineering ADCET,
Ashta, Shivaji University Kolhapur, India

Abstract: Cloud computing using different services like server storages, software development platforms which are kept on the internet to update, solves and process data. To place the different applications without any addition cloud computing provides no of processing and storage capability to the user. The huge collection of intermediate datasets will be generated and it saves the cost. But there is a problem to preserve the privacy of such intermediate datasets. The adversaries check the sensitive data by studying multiple no of intermediate datasets. In existing application, all datasets are encrypted in cloud. But to encrypt all datasets are not enough and price oriented that's why, it takes more money and time for encryption and decryption of datasets. So proposed a revelation based application to recognize which middle dataset need to be encrypted and which do not. It takes less privacy preserving cost using different algorithms. The module to generate signature of stored data which is stored on the server .and compare with original one. Keep one copy of that data at our side and compare with original one. If it matches then say our data is safe and if it not matches then say data is corrupted .To avoid high recompilation they store some valuable intermediate dataset which is considered.

Keywords: Privacy preserving, Cloud Computing, Data Storage privacy, Intermediate datasets and Privacy upper bound.

1. Introduction

Cloud Computing contains the different networks, interfaces, storages, services those are collected to provide as a part of computing a service. It provides the storage services and application services over the internet like different components or whole platform related to requirements of user. Through this application users stores numerous set of information on cloud which provides greater flexibility, storage and computation of data but these applications can be processed, large set of processing data which is created. Cloud computing has information based service to obtain computer resources and network storage space. Cloud also says as "give-as-you-go" fashion means that cloud service provider to purchase required storage space if technical needs change. In cloud application, when processing on the original datasets the intermediate set of data has to be generated. Users in cloud environment have more storage space and computation. So, it store price able dataset on a cloud to reduce regenerating cost of datasets. Storing intermediate dataset in cloud shows the malicious surface area.

The existing application includes anonymization and encryption techniques for preserving the privacy of intermediate datasets. Traditional application that hide all intermediate datasets with the use of elgamal encryption technique. But such datasets are rarely taken by no of users and collected frequently. It encrypts or decrypts data repeatedly.

To overcome such traditional problem we encrypt part of intermediate data sets except all datasets for decreasing privacy preserving cost. So proposed an application to recognize which middle datasets required to encrypt and which do not. To preserve multiple datasets privacy, it annonimize all datasets first then encrypt them in cloud.

2. Related Work

The Authors[1] proposed an application to recognize which part of datasets required to encrypt so that preserving cost can be saved. The research through understand how to protect privacy in Cloud area.

The Authors[3] converting data into secret code which is described by old application to show the personally identifying information in cloud. In many applications, it is necessary to convert data sets again. Secret codes are normally attached to some methods to gain privacy protection, cost reduction, high data usability.

The Authors[2] gave a cloud computing view. Such utility has the power to transform big part of industry. It refers to the application provided services on the different network centers.

The Authors[4] evaluate that the annonimization effects on the future data classification which preserves the privacy of data.

The Authors[5] tells that cloud computing is a current generation architecture of IT Enterprise. They first discover the queries and protection problems and check the scheme. The Authors[6] to improve the application data confidentiality which is stored on third party cloud computing. It proposes to encrypt and recognize all functionally encryptable data. The Authors[8] assign secret keys to specific part of data to decrease planning of keys and provides the device data access. The need of retaining middle set of data on environment is widely understood, but the protection reasons in research solved by such sets of information.

3. Need and Objective of Proposed Work

When we check current market scenario that lots of organizations are moving towards cloud and it is very cost effective to store large amount of data on cloud, at that time we feel that we can provide a solution for cost effective data store on server without losing our important data. When we go through the paper the "A Privacy Leakage Upper bound" we get that whatever the solution we are going to give that save the cost of storing all data without encrypting un-necessary data.The objectives of proposed work is as follows, The new application gives the storage data security.

- The system reduces the cost of maintaining Intermediate datasets.
- The system preserving the privacy of data.
- The model verifies stored data.

4. Research Methodology

4.1 The following modules are included in the architecture,

4.1.1 Data Storage Privacy Module:

The model stores the patient original data. After deleting personally identifiable information data user's process or access part of original data. Users like research center, government, adversaries, and pharmaceutical company are included. The Data storage model views whole information about the patient. The new application mix up secret code and data files storing to gain protection in data saving which is divided.

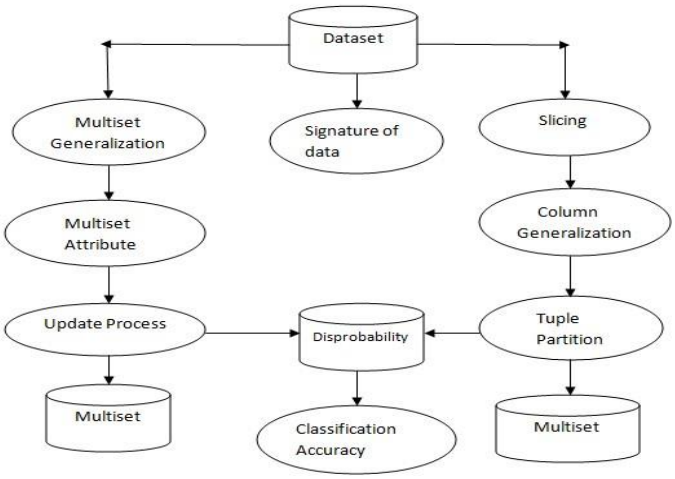

Figure 1. System Architecture

4.1.2 Privacy Preserving Module:

Every organization needs particular data about patient so, to preserve the privacy of multiple datasets used the security layers. There are five security layers. Each layer contains different information about different intermediate datasets which are part of original dataset. The layers protect the personal data. So it preserves the datasets protection. The learning to achieve new things algorithm is used for privacy preserving.

4.1.3 Intermediate Dataset Module:

To reuse data and save the cost middle coming datasets made during data processing .The model first provides user registration with user type. After registration user login the account and searches the patient agewise and diseasewise based on data user's type.

4.1.4 Privacy Upper Bound Module:

Privacy upper bound module to select the required data from intermediate datasets. This module encrypts part of original data not whole data .The remaining data will be in decrypted form. So it saves the time and cost of privacy preserving.

4.1.5 Data Signature Module:

Data signature module to generate signature of data which is stored on the server. We can generate signature of that uploaded data and compare with original signature. The new application to match whether it is correct or not. If it matches then we can say our data is safe but if there is any change then we can claim that data is corrupted.

4.2 Algorithms

4.2.1 Heuristic Algorithm-

Disc: Repetition shows the middle dataset which need to encode a message, gaining less level privacy maintaining price.
Input: Original Dataset and middle dataset with properties–Frequency, Size, Privacy requirement threshold Privacy leakage.
Output: New intermediate dataset that need to be encrypted.
Step1: initialize variables
Step2: Define a priority queue: PQueue
Create a root of SIT (SN) with parameters –privacy leakage and encryption cost
Calculate privacy leakage and encryption cost
Check whether encryption of dataset (ED[i]) is optimal, If yes solution is found, and go to step 3.

If No, Get intermediate dataset which are not encrypted, Sort dataset by ascending order and calculate its Privacy leakage and Encryption Cost Construct new dataset and add it to PQueue.
Step3: Mark dataset as encrypted & Show data from selected dataset

4.2.2 Elgamal encryption algorithm

The Elgamal encryption algorithm used to encrypt the data fast and easily. This is an equal basic encoding a message for main secret cryptography. It is based on main common network. The system written an extra row of security by unequally encoding a message already used for unequal message encoding. In 1985 TaherElgamal was founded by Elgamal algorithm .There are two components: the Basic Alternator, the elgamal encryption.

The Basic Alternator arrangement The Basic Alternator performs like below:
Person X alternate's powerful information of a repeating set G of rule q along with generator g.
The properties of set G are as follows:

- Person X accepts an inconstantly against {1… q-1}. Person X counts p: $=g^a$.
- Person X declares p, with powerful information of G, q, g, as its Basic key.
- Person X contain a as its secret key, that put as private key.

Encryption Arrangement

The Elgamal encryption performs like below:
To encode a message w to Person X under her Basic key {G, q, g, p}.

- Person Y accepts inconsistent b against {1… q-1}, later counts c1:$=g^b$. Person Y counts the common private v: $=p^b$: $=g^ab$. Person Y designs his message w towards a component w' of G.
- Person Y counts c_2:=w'.v.
- Person Y writtens ciphertext (c_1, c_2) = (g^b, w'.g^ab) to Person X.

4.2.3 Frequent Pattern Algorithm

Input: An Enterprise database and smallest base origin.
Output: frequent pattern tree, the frequent pattern tree of database.

a. First search the enterprise database. Save the frequent item set F and each frequent items base. Group F in base-downward line as FList, the list of frequent items.
b. Construct the origin of frequent pattern tree , T, and tag it as "invalid". For each Enterprise transaction in database works as follow:

- The frequent items in Transaction should be selected and save them based on the line of FList. Let the sorted frequent-item list in Trans be [p | P], where a is the first element and A is the remaining list. Call insert tree ([a | A], T).

- The operation insert tree ([a | A], T) is shown as follows. If T has a child C Proves that C.item-name = a.item-name, so increment C's count by 1; otherwise make a new node C, with its count initialized to 1, its parent line linked to T, and its format. If A is nonempty, call insert tree (A, C) regardfully.

5. Results and Discussion

Analyzed performance of Encryption for new set of layers with existing search based Privacy preserving of intermediate Datasets Figure 5.1, 5.2 and 5.3 shows the evolution of encryption time, encrypted dataset size and Cloud Storage and Encryption cost for new set of layers in existing system and proposed system.

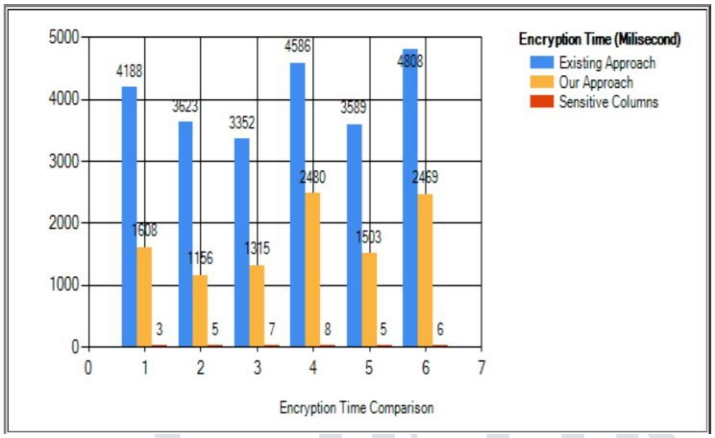

Figure 5.1. Encryption Time Comparison

This experiment aims to examine the difference between the Encryption time for EA and OA, respectively. The SA shows the sensitive columns. The Encryption time for EA, OA and SA are illustrated in Figure 5.1. Some observations can be drawn from the experimental results. The existing system takes more time to encrypt the data because it encrypts whole data in the database. But our system takes less time to encrypt data because it encrypts required data except whole data. Based on elgamal encryption we found less time in milliseconds to encrypt data and also it saves the time.

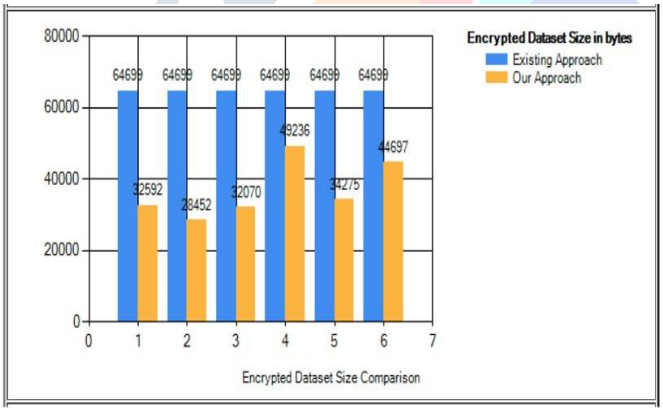

Figure 5.2. Encrypted dataset size comparison

This experiment aims to examine the difference between the Encryption time for EA and OA, respectively. The Encryption time for EA and OA are illustrated in Fig5.2.In this comparison the existing system takes extra space to store large no of datasets. Our approach takes smaller size to store the datasets. The data stores in bytes. If load is less then it works faster so easily we can handle the process. By comparing both the approaches the new approach is best.

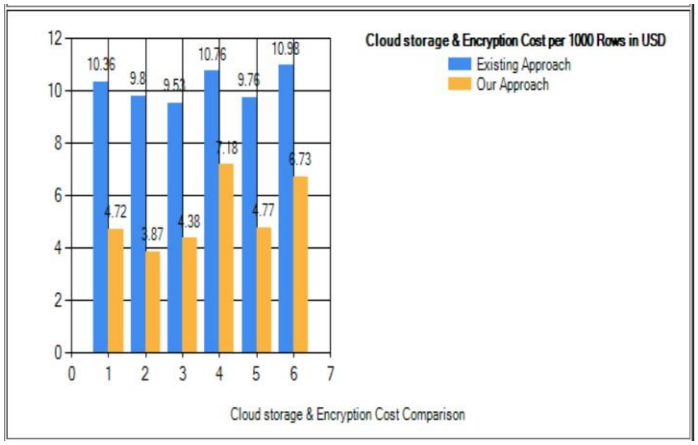

Fig5.3.Cloud Storage & Encryption Cost Comparison

This experiment aims to examine the difference between the Encryption time for EA and OA, respectively. The Encryption time for EA and OA are illustrated in Fig5.3.In the existing approach the large number of data is stored on server so it needs more space in cloud storage. For more space we need more memory space so it is costly to store all data in the cloud but in our approach we are storing only required data so the cost is also less. The cost is compared in US dollar the experimental setup shows cloud storage and encryption cost is appropriate in new approach. As per comparison shows that it takes less time to encrypt data,stores data in smaller size and it require minimum cost.

Acknowledgments

I would like to express my deep sense of gratitude towards my guide and co-guide for her invaluable help and guidance for the dissertation work. I am highly indebted to her for constantly encouraging me by giving critics on my work.

References

[1] Xuyun Zhang, Chang Liu, Surya Nepal, Suraj Pandey and Jinjun Chen, "A Privacy Leakage Upper Bound Constraint-Based Approach for Cost-Effective Privacy Preserving of Intermediate Data Sets in Cloud", IEEE TRANSACTIONS ON PARALLELAND DISTRIBUTED SYSTEMS, VOL. 24, NO. 6, JUNE 2013.

[2] M. Armrest, A. Fox, R. Griffith, A.D. Joseph, R. Katz, A. Konwinski, G. Lee , A View of Cloud Computing, vol. 53, no. 4, pp. 50-58, 2010.

[3] D. Yuan, Y. Yang, X. Liu, and J. Chen, "On-Demand Minimum Cost Benchmarking for Intermediate Data Set Storage in Scientific Cloud Workflow Systems," J. Parallel Distributed Computing, vol. 71, no. 2, pp. 316-332, 2011.

[4] B.C.M. Fung, K. Wang, and P.S. Yu, Anonymizing Classification Data for Privacy Preservation, vol. 19, no. 5, pp. 711-725, May 2007.

[5] Qian Wang, Cong Wang , Jin Li1, Kui Ren1, and Wenjing Lou , Enabling Public Verifiability and Data Dynamics for Storage Security in Cloud Computing.

[6] K.P.N. Puttaswamy, C. Kruegel, and B.Y. Zhao, Silverline: Toward Data Confidentiality in Storage-Intensive Cloud Applications, Proc. Second ACM Symp. Cloud Computing, 2011.

[7] E.T. Jaynes, "Information Theory and Statistical Mechanics," Physical Rev., vol. 106, no. 4, pp. 620-630, 1957.

[8] S. S. Sayyad, D. B. Kulkarni, "An Encrypted Neural Network Learning to Build Safe Trained Model", International Journal of Computer Sciences and Engineering, Vol.06, Issue.01, pp.32-36, 2018.

[9] Mr. S. S. Sayyad, Dr. P.J. Kulkarni, "Privacy Preserving Back Propagation Algorithm for Distributed Neural Network Learning", International Journal of Scientific and Research Publications, Volume 2, Issue 3, March 2012.

[10] https://en.m.wikipedia.org/wiki/ElGamal_Encryption.

[11] https://en.m.wikibooks.org/wiki/Algorithm.

[12] In_R/Frequent_Pattern/The_FPGrowth_Algorithm#The_Algorithm.