ISSN 2348 – 7968

Cost-effective Privacy Preserving of Intermediate Datasets in Cloudusing Reverse Circle Cipher

SuyashMendhe,AkshayGunjal , AmolDarekar ,KapilTadewade and Mrs.SuvarnaKadam

Student, D.Y. Patil College of Engineering, Akurdi.
suyashmendhe3399@gmail.com
Student, D.Y. Patil College of Engineering, Akurdi.
akshaygunjal9369@gmail.com
Student, D.Y. Patil College of Engineering, Akurdi.
darekaramol21@gmail.com
Student, D.Y. Patil College of Engineering, Akurdi.
kapiltadewade@yahoo.com
Assistant Professor, D.Y. Patil College of Engineering, Akurdi.
suvarna.kadam@gmail.com

Abstract

Cloud Computing, recently emerged computing paradigm transforms the perception about infrastructure investment in IT sector. Cloud computing has an edge over grid computing, utility computing and autonomic computing and thereby causes rapid transition towards clouds. Although it has some problems about security of data, at time of processing any data intensive application huge amount of intermediate datasets are generated and they are often stored for future purpose instead of generating them. Privacy maintenance of such intermediate dataset is a challenge because data owner can easily access item. Existing approach uses encryption of all such intermediate dataset to provide privacy, but this consumes takes lot of time and cost. To make this cost-effective, we have testedthe upper bound constrain which will encrypt only part of intermediate datasets rather than whole, indirectly reducing privacy preservation cost and satisfying privacy requirement of users.

1. INTRODUCTION

With Advancement in IT over last 2 decades, computing became the 5th utility of day to day life. Cloud computing is a way to get massive computational capabilities and storage area for users. Clients can store huge amount of data on cloud. Some data may contain sensitive information, some may not. But to store sensitive information in cloud is more difficult problem. Such information sharing is subject to privacy constraints of data subjects and also data confidentiality of data provider.

Privacy concern arises due to maintaining the intermediate datasets in cloud is serious concern but given less attention. From economical point of view storage and computational services are equivalent they are charged according to their usage. Hence cloud user should take wise decision about which intermediate datasets should store selectively while processing original datasets in data intensive application to cut down the total expense by avoiding repeating computation to obtain these datasets. It is more common that user may alter the intermediate datasets, sharing it with other or analyses it periodically. So generally it is sense that intermediate datasets are referred to other intermediate datasets and resultant datasets. By storing such intermediate datasets increases chances of attack so the privacy of data holder is in danger. Common multiple parties access and process such intermediate dataset without control of original data holder. So adversary can collect these different dataset together and perceiving threat of disclosing privacy and causes economical loss or reputation in society.

Existing privacy preservation approaches includes two main techniques that are encryption and anonymization. Current research mainly includes the encryption of all the datasets. However, processing of such encrypted datasets effectively and efficiently is quite challenging. But recent research enables us to perform computation on encrypted data which theoretically feels good but practically it is very expensive. In some cases, summarized information has to be provided to user. So anonymization is preferred over the encryption. Current techniques hold good for single datasets but privacy preservation of multiple datasets is still an issue. Thus, it is advised that first do anonymization of all datasets and then encrypt them before deploying them on cloud. Content wise intermediate datasets are very vast. Hence encrypting all intermediate dataset causes low efficiency and additional overhead of encryption when accessed frequently. So we have proposed to encrypt selective intermediate dataset for cut down the cost further.

To identify which intermediate datasets to be encrypted and which not, we propose new approach to satisfy data holder privacy constrain. How the intermediate datasets are generated from their parent dataset is analyzed and then these relationships are molded to dependency graph forming tree structure. Determining the value of privacy leakage through multiple dataset we use upper bound constraint to limit privacy disclosure problem is divided into number of sub problem by decomposing privacy leakage constraint. This leads to practical solution that identifies which intermediate dataset to be encrypted. Cost of the encryption is reduced when we conduct our experiment on real world and extensive datasets.

2. LITERATURE SURVEY

Encryption is the one of the most popular technique utilized for security of data in cloud. There are many approaches used for encryption of data.

Benjamin c m [1] proposed privacy preservation of data publication. It provides methods and tools for publishing useful information while preserving data privacy. It is mostly useful in data mining research community. But it losses valuable information, increases complexity and cost of publishing.

Another approach for encryption is anonymization where some data is anonymized before publishing into cloud. This technique is similar but not beneficial for large number of service provider. Encryption and fragmentation approach[5] pair encryption together with fragmentation. In this method fragmentation is done in order to break the link between two sensitive datasets and then encrypting necessary data. Breaking of information into small fragments will break association between sensitive data. User can access data through encryption key.

Airavat[6] is a novel integration of mandatory access control and differential privacy. Mandatory access control by which the operating system constraints the ability of a subject or initiator to access or generally perform some sort of operation on an object or target. Differential privacy provides means for maximizing accuracy of queries from statistical dataset and also minimizes the chance of identifying its records. Limitation of airavat is that it fails to confine every computation performed by untrusted code. Silverline[7] is a set of tools that automatically identifies all functionally encrypted data in a cloud system. Then it assign encryption key to specific data subset to minimize complexity of key management. As all data on the cloud is not encrypted, so cloud may be vulnerable to attacks. Sedic[8] gives a solution to the privacy by keeping the computation on the private data within organizations private cloud and moving the rest to public cloud. Here sensitive data always remain on the private cloud. This technique has overhead of transferring data on public and private cloud.

3. REVERSE CIRCLE CIPHER

For personal security point of view block cipher techniques such as DES(Data Encryption Standard) and AES(Advanced Encryption Standard) are used. But multiple passes over each block and matching them is ineffective for real time data transfer. Similarly Diffiehellman and RSA(Rivest Shamir Adlemsan) require large number of bits resulting into increment of time and space complexity.

Reverse circle cipher uses ‘Circular substitution‘ and ‘reversal transposition’. This encryption technique uses an arbitrarily variable key length which may be equal to length of plaintext or can be very small. RCC neither works in bit level nor it manipulate the bytes orientation. Rather it changes ASCII value of text. Confusion created by circular substitution by replacing the original text by the equivalent text obtained by adding a user given string to original ASCII value of string resulting to newly generated ASCII value. When index position of string reaches maximum point, the position restart to start index until whole conversion of plain to cipher text takes place. This phenomenon refers to circular key mechanism.

Reversal transposition implements diffusion by buffering a certain amount of character of plaintext and writing in reverse order into cipher file the amount of characters picked in buffer is called as reverse length.

Decryption process is the inversion of encryption process.

3.1 Encryption Algorithm

```
Step 0: Start
Step 1: Get Input String S
Step 2 : Initialize a String ENC as empty
Step 3: Divide the string S in  N blocks of size    10 characters
Step 4: for I =1 to N
Step 5: Let String BS =10 character of each block
Step 6: rotate block with I characters in clock     wise
Step 7: for  j=1 to 10
Step 8: substitute each character
Step 9: With special character
Step 10: End of inner for
Step 11: ENC=ENC+BS
Step 12:End of Outer for
Step 13: Stop
```

3.2 Decryption Algorithm

```
Step 0: Start
Step 1: Get Input String S
Step 2 : Initialize a String DCR as empty
Step 3: Divide the string S in  N blocks of size 10 characters
Step 4: for I =1 to N
Step 5: Let String BS =10 character of each block
Step 6: rotate block with I characters in anti clockwise
Step 7: for j=1 to 10
Step 8: substitute each character
Step 9: With special character
Step 10: End of inner for
Step 11: DCR=DCR+BS
Step 12: End of Outer for
Step 13: Stop
```

It is only optimizing performance of data in transit but also provide adequate level of security of data. It is proved very difficult to break by Ebenezer Issac[9].

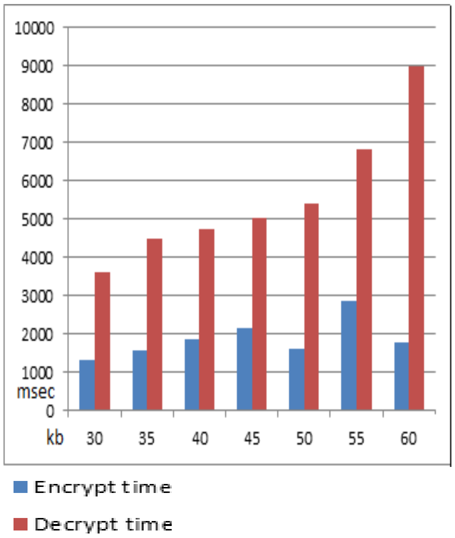

Fig1. Encrypt time and decrypt time for txt file

Fig1.Shows that reverse circle cipher is more efficient than other algorithms.

4. EXPERIMENTAL PROCESS

To show the effectiveness of proposed system some experiments are conducted on java based windows machine. To measure the performance of the system we set the bench mark by detection accuracy of the intermediate data sets in cloud. To determine the performance of the system, we examined how many relevant intermediate data sets are identified based on the upperbound constant in cloud.

To measure this precision and recall are considering as the best measuring techniques. So precision can be defined as the ratio of the number of relevant intermediate data sets are identified to the total number of irrelevant and relevant intermediate data sets are identified. It is usually expressed as a percentage. This gives the information about the relative effectiveness of the system. Whereas Recall is the ratio of the number of relevant intermediate data sets are identified to the total number of relevant intermediate data sets are identified. It is usually expressed as a percentage. This gives the information about the absolute accuracy of the system.

The advantage of having the two for measures like precision and recall is that one is more important than the other in many circumstances.

For more clarity let we assign

- A = The number of relevant intermediate data sets are identified,
- B = The number of relevant intermediate data sets are not identified, and
- C = The number of irrelevant intermediate data sets are identified .

So, Precision = (A/ (A+ C))*100

And Recall = (A/ (A+ B))*100

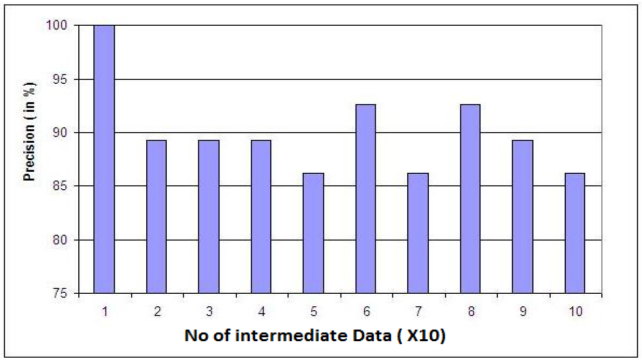

Fig.2. Average precision of the proposed approach

In Fig. 2, we observe that the tendency of average precision for the identified intermediate datasets are high compared to other systems.

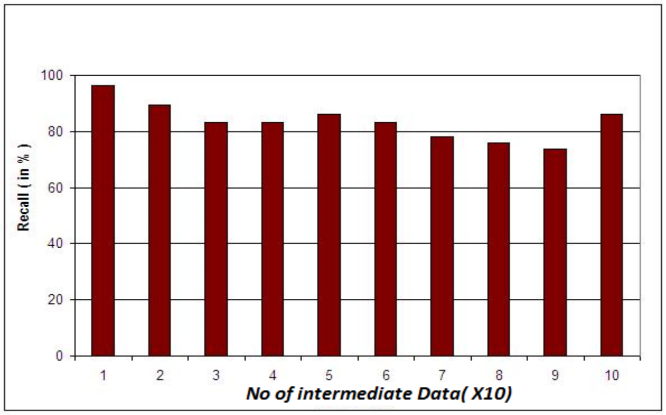

Fig. 3. Average Recall of the proposed approach

In Fig. 3, we observe that the tendency of average Recall for the identified intermediate datasets is high compared to other system. So this shows that our proposed system is achieving high accuracy than any other method.

5. CONCLUSION AND FUTURE SCOPE

This approach identifies which part of intermediate data sets needs to be encrypted while the rest does not, in order to save the privacy preserving cost. Data sets and larger extensive data sets have been shown the cost of preserving privacy in cloud can be reduced significantly with our approach over existing ones where all data sets are encrypted.

Intermediate data set management is becoming an important research area in cloud. Privacy preserving for intermediate data sets is one of important yet challenging research issues, and needs intensive investigation. Approach can be further evolved to provide security for multimedia such as images using suitable encryption algorithm. Tree construction in upperbound can be further transformed to graph structure for identifying which data sets need to be encrypted. Privacy preservation matrix can be taken together with storage and computation. Various cloud service providers can use this approach as application to provide security.

ISSN 2348 – 7968

6. REFERENCES

[1] 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.

[2] Wang J, Zhao Y et al. (2009). Providing Privacy Preserving 1. In cloud computing, International Conference on Test and Measurement, vol 2, 213–216.

[3] L. Sweeney, "k-anonymity: a model for protecting privacy", International Journal on Uncertainty, Fuzziness and Knowledgebased Systems, 2014, pp. 557-570.

[4] A.Machanavajjhala, J.Gehrke, and D.Kifer, et al, "ℓ- diversity: Privacy beyond k-anonymity", In Proc. Of ICDE, Apr.2014.

[5] 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.

[6] Roy, S.T.V. Setty, A. Kilzer, V. Shmatikov, and E. Witchel, "Airavat: Security and Privacy for Mapreduce," Proc. Seventh USENIX Conf. Networked Systems Design and Implementation (NSDI '10), p. 20, 2011.

[7] 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 (SoCC '11), 2011.

[8] K. Zhang, X. Zhou, Y. Chen, X. Wang, and Y. Ruan, "Sedic: Privacy- Aware Data Intensive Computing on Hybrid Clouds,"

[9] Ebenezer R.H.P. Isaac, Joseph H.R. Isaac and J. Visumathi "Reverse Circle Cipher for Personal and Network Security" Jeppiaar Engineering College Chennai, Tamil Nadu, India ebeisaac@gmail.com 2014