International Journal of Scientific Engineering and Applied Science (IJSEAS) - Volume-1, Issue-3, June 2015
ISSN: 2395-3470
www.ijseas.com

Extracting Meta data using data de-duplication in hybrid cloud

V Ganapathy Sundaram[1], P Saravanan M.E[2]
[1] PG Student, [2]Assistant Professor, Department of Computer Science and Engineering,
PTR College of Engineering and Technology, Tamilnadu, (India)

ABSTRACT

The cloud data base service is a boon to the global technology which support various internet based application. But the serious issue behind this approach is the data confidentiality which threats the user's dealing with the sensitive data's. In order to overcome this problem, we propose a novel approach of adaptive encryption of cloud databases. That ensures the data security along with cloud database structure flexibility at design time. We have taken (DBaas) database-as- a- service and compute our proposed cost estimation model to overcome the several security challenges according to the real-time scenario. Our proposed scheme uses meta-data with advanced encryption term according variable price and data load during a medium period of time. To make the process more prominent we also dealing with data de-duplication so the work load minimize and exhibits an effective overall result.

Index Term: Cloud database, Data de-duplication, confidentiality, Adaptive encryption, Cost estimation model

INTRODUCTION

According to today's globalization cloud computing allows various facilities for the users in several aspects such as Naas, DBaas, Paas, Saas and Iaas. In the current world cloud providers facilities storage space and vast parallel computing sources in a minimum cost. Cloud computing enables various security aspects in order to provide the client users for accessing the data and rights on the stored data. The important issue on the cloud storage services is managing of large amount of data. A scalable data management is possible by a well known data computing application such de-duplication. The reason for this massive effectiveness is problem of repeated data occupies the large amount of storage space which affects the overall cloud computing process of a network. On discussing about de-duplication, it is an improvised data compression technique avoids repeated data to be stored. It not only occupies the spaces, by processing the overall data during transmission it increase the bandwidth as well as overload package. Having content with same data, de-duplication enables the original data by keeping only one physical copy instead of multiple ones by referring the repeated data to that copy. De-duplication can be applicable to place on both levels such as file level or block level. On discussing about file level it checks the repeated data within that file. To the other aspect on block level it avoids data copies which are present at the non-identical files. Data de-duplication provides not only space but also security and privacy on the user's sensitive data which was the serious issue on the client aspect. On de-duplication data privacy must be secured from the malicious attackers within as well as outside the network.

International Journal of Scientific Engineering and Applied Science (IJSEAS) - Volume-1, Issue-3, June 2015
ISSN: 2395-3470
www.ijseas.com

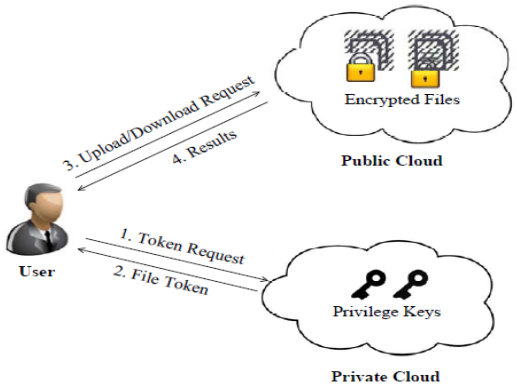

Fig 1: Architecture of Authorized De-duplication

To make this privilege encryption concept is evolved there are various traditional encryption techniques are followed to ensure the data confidentiality in a network. Because most of the existing methods requires the data owner to encrypt their data with own keys. But the problem is unexpectedly the data from different owners enables multiple cipher text that makes the de-duplication impractical. The next this is to secure those data from the unauthorized users to achieve this separate protocol is required to owns the file if there any duplicated data's are found. By the proof the applicable file is provided from the server instead of uploading the same one. That file can be downloaded by the authorized user and can be decrypted only with the convergent keys otherwise the data cannot decrypt which proves the data security to the owner. By doing this methodology it proven that cloud makes the de-duplication and convergent prevent the sensitive data from the unauthorized users.

RELATED WORKS

The increasing popularity and cloud database advantage makes the researchers to contribute their works in improving the overall functionality. Initially the size of the database is taken into the concern before it uploaded to the cloud for this de-duplication approach is computed and then the information confidentiality is discussed at [1] [2] over a long medium time [3] [4]. Previously various researches were discussed about the Database as a Service paradigm (DBaaS) and its security issues at the tenant's point [5]. Some works execute SQL operation on encrypted data but those suffers from limitation of effective performance [6] [7] [8]. In these the SQL operations which are statically determines the set of queries during the design time [9] [10] but it assumed on LAN scenario which does not have a network latency. In the next stage they uses the Meta data and encrypt those metadata using data aggregation technique but the plain mata data has the possibility of leaking sensible data. As well as the aggregation model leads to various network overheads. According to the fully homomorphic encryption [11] it satisfied the execution of encrypted data but it suffers from huge computational costs. The other encryption algorithm which have minimum computational cost that SQL operators by means of order comparison command but not appropriate to search operators [12] [13] [14]. This drawback is due to inadequate knowledge of data administrator regarding the design time at which database operations will be required over each database column. In paper [15] integrate adaptive encryption schemes with a proxy free architecture is discussed but the data confidentiality and its cost is at the concern of cloud tenant organizations. The costs of cloud computing from a provider's perspective are discussed by the author at [16] [17] but these are unfamiliar due to its drawbacks like servers, power consumption, and infrastructures which are not supports to the analytical cost estimation model. According to the

CloudSim [18] it helps to maintain and estimate multiple cloud servers alternatively but it does not effective to medium-term period as the work load varies the cloud prices. On dealing with the meta-data the cloud providers were defined semi-honest or honest-but curious
[19], they do not modify the data by using the SQL operations they could be interested in accessing tenant's information stored in the cloud database. In which the cost estimation model is evolved by the work analysis of most popular database services such as Amazon Relational Database Service [20], EnterpriseDB [21], Windows Azure SQL Database [22], and Rackspace Cloud Database [23] along with Double Extra Large from Amazon RDS [24].

EXISTING SYSTEM

The existing system uses traditional de-duplication approach which is lead to have large amount of data. It does not support differential authorization check which is considered as a main task in most of the applications. Generally earlier approaches are based on the convergent encryption but those are not effective duplicate check which the increase data load by means of repeated data's. In those approaches the file transferred and stored in cloud without any authentication that maximize the problem of data privacy. There is no prominent guarantee on dealing with the sensitive records however it supports adaptive encryption scheme. It uses SQL operations and queries for extracting the data which has minimum level of scalability. The existing system deals with the plain text metadata may leak sensitive information. Moreover the existing implementations are affected by huge computational costs to the extent that the execution of SQL operations.

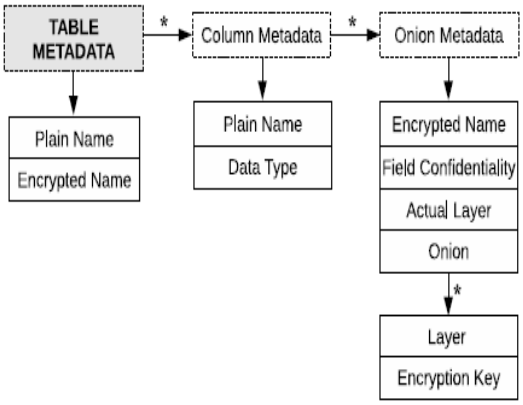

Fig 2: Metadata Structure

PROPOSED SYSTEM

With the motto of achieving data confidentiality we designed a proposed architecture which supports adaptive encryption technique that encrypts the meta data before stored in the cloud. It supports the SQL operations like SELECT, INSERT, UPDATE and DELETE by means of encrypted database.

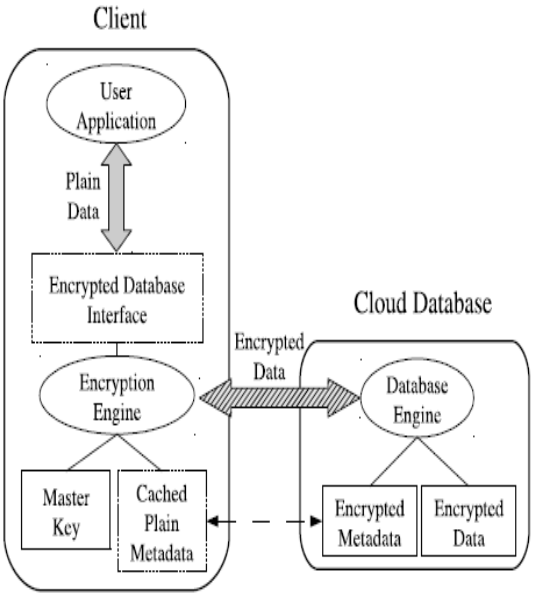

Fig 4: The proposed Architecture

The proposed architecture undergoes it process by means five set of information's such as

- Plain data which holds set of tenant information's
- Encrypted data that is to be stored in the cloud database
- Plain meta data which contains the valid information that is used perform SQL operation
- Encrypted metadata is nothing but a plain data in encrypted format to ensure the confidentiality
- Finally master key is the encrypted key so that it is distributed to every legitimate user in the cloud.

Based on this information' the whole process is happened as shown in the fig 4 with the help of adaptive encryption scheme and the working mechanism of this scheme is explained below;

- Random (Rand): It is consider as most secure encryption method which does not reveal the original plain is only applied for information retrieval as it suffers from major drawback that does not supports SQL operations.
- Deterministic (Det): In which the plain text is secured by encrypting the data and supports equality operator
- Order Preserving Encryption (Ope): It safeguards the encrypted values numerical order of the original unencrypted data and also supports SQL comparison operators such as =, <, ≤,>, ≥.
- Homomorphic Sum (Sum): It carries the function of sum operator by handling the multiplication encrypted integers is equal to the sum of plaintext integers.
- Search (Search: It is nothing but a LIKE operator help to check the full string.
- Plain: It handles all SQL operation on means of non confidential data as it does not applicable on encrypted data.

To be more efficient our proposed scheme also supports cost estimation model by means of three parameters such as time, pricing and usage. In which the time gives the details about intervals for which the tenant service to performed. Next the pricing is calculated by the resource usage and subscription of cloud provider then the usage is total amount of resources used by the tenant. The overall calculation is defined by;

$$\cos t = f(Time, Usage, \Pr ice)$$

IMPLEMENTATION DESIGN

A software prototype is designed using dot net framework to implement our proposed architecture in an efficient manner. In which the initial stage is preparing the data base to be processed in the application and the data base hold the column *name*, *data type* and *confidentiality parameters along with the tenant policy with* strongest encryption algorithm as discussed in the above section.

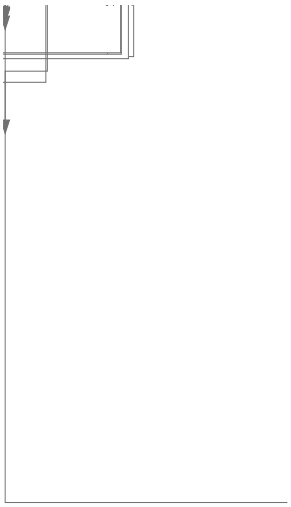

Fig 5: Implementation design

As shown in the figure a intially a login module is create to ensure that only authorized user can acess the process. Based on that a valid user enters into the application processing the created database and which undergoes deduplication by means of token generation. Next the duplicate dataare removed and using the algorithm th data along with the metadata is encrypted and stored in the cloud file . The encrpyted key is shared to the valid user for retriving the orginal data even by means of SQL operations as it supports on the encrypted data as discussed in the above section.

Result and discussion:

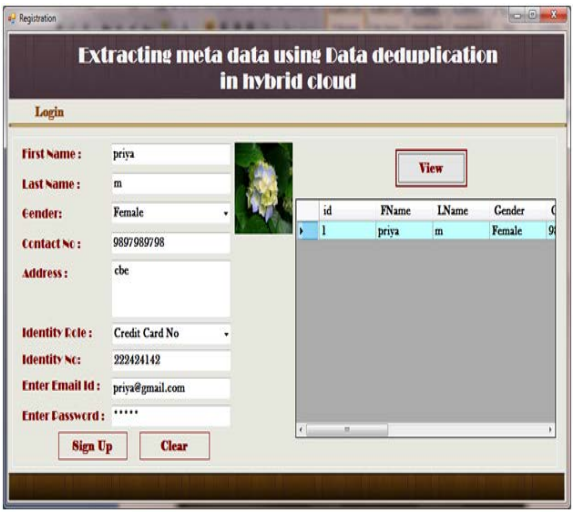

Fig 6: User authetication

International Journal of Scientific Engineering and Applied Science (IJSEAS) - Volume-1, Issue-3, June 2015
ISSN: 2395-3470
www.ijseas.com

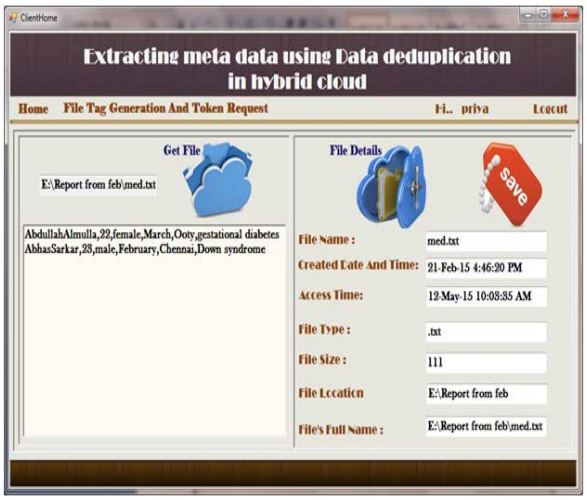

Fig 7: Flag setting

From the figure 6 & 7 a user entering the details to get acess in managing the databased in the cloud. In which the created data base is to be uploaded from the location. It haves all the details like meta data data , file location, file size, created data along with time. The tag sets the data ID to be processed each one as unique in the cloud data base server.

Fig8 : Encrypted data

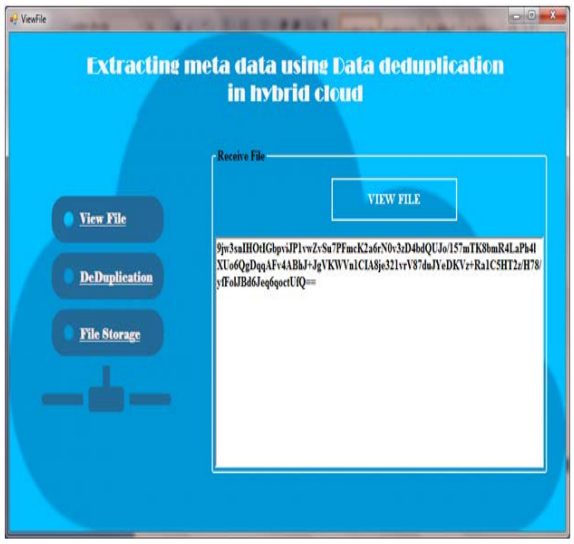

Fig 9: Duplication checking

The result from fig 8 & 9 shows the time , price and usage of the processes in order to progress on cost estimation model along with the further details the plain data is encrypted with the key and stored in the cloud data base.Before the data is under goes duplication check by means of token after confirming no duplication the viewed file is stored in the prescribed location those details were shared to the appropriate data owner in a secured manner.

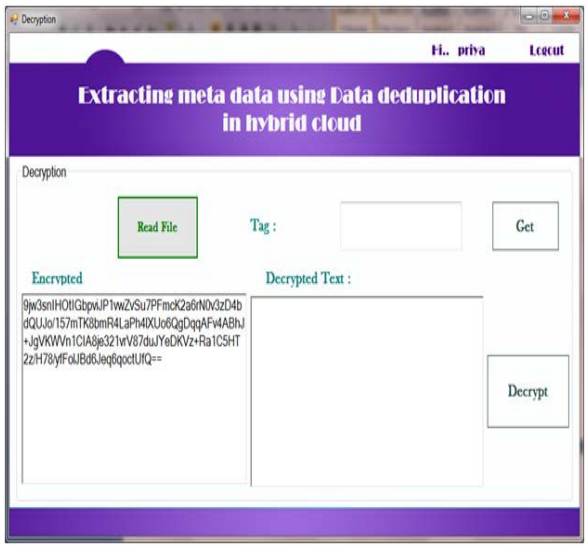

Fig10: Encrypted file extraction from cloud

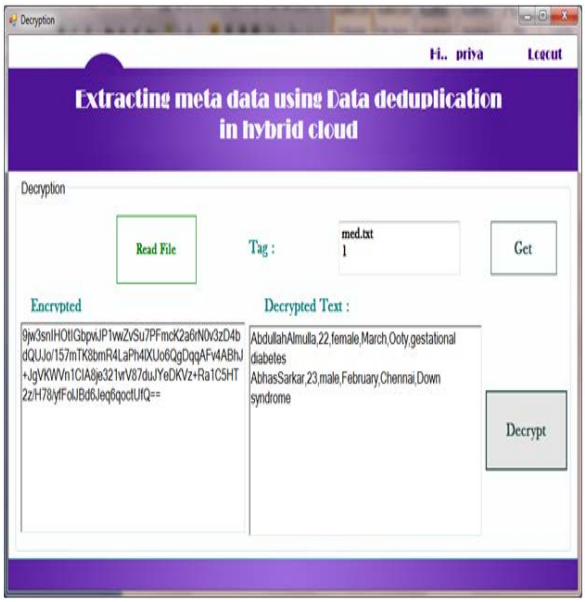

Fig 11: Decrypted data

As in fig 10 & 11 shows the encrypted data which was stored in the cloud location by using the tag id and then the original data is decrypted by sharing encrypted key with original user as explained in the proposed scheme. The effective of the adaptive encryption scheme and data confidentially is shown in the fig 11. From those result it justified that the proposed system is far better than tradition method in maintaining the data de-duplication along with data privacy in cloud database server applications.

CONCLUSION

In this paper we have discussed about duplicate data and confidentiality with cost analysis which are considered as recent issues in cloud database services. Our proposed novel cloud database architecture that uses adaptive encryption techniques investigates using tenant's information in improving the feasibility and performance through a software prototype in a prominent manner. And the extracted result shows the hiding of sensitive data from the unauthorized users by encrypting the meta data using adaptive encryption. Thus the loaded data without duplication and secured manner implies the proposed work as better according to the real time scenario and further investigation is carried out in future in different threat model hypotheses.

REFERENCE

1) 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 5th utility," Future Generation Computer Systems, vol. 25, no. 6, pp. 599–616, 2009.

2) T. Mather, S. Kumaraswamy, and S. Latif, Cloud security and privacy: an enterprise perspective on risks and compliance. O'Reilly Media, Incorporated, 2009.

3) H.-L. Truong and S. Dustdar, "Composable cost estimation and monitoring for computational applications in cloud computing environments," Procedia Computer Science, vol. 1, no. 1, pp. 2175 – 2184, 2010, iCCS 2010.

4) E. Deelman, G. Singh, M. Livny, B. Berriman, and J. Good, "The cost of doing science on the cloud: the montage example," in Proc. 2008 ACM/IEEE Conf. Supercomputing, ser. SC '08. Piscataway, NJ, USA: IEEE Press, 2008, pp. 50:1–50:12.

5) H. Hacig¨um¨us¸, B. Iyer, and S. Mehrotra, “Providing database as a service,” in Proc. 18th IEEE Int’l Conf. Data Engineering, Feb. 2002.

6) G. Wang, Q. Liu, and J. Wu, “Hierarchical attribute-based encryption for fine-grained access control in cloud storage services,” in Proc. 17th ACM Conf. Computer and communications security. ACM, 2010, pp. 735–737.

7) Google, “Google Cloud Platform Storage with server-side encryption,” http://googlecloudplatform.blogspot.it/2013/08/ google-cloud-storage-now-provides.html, Mar. 2014.

8) H. Hacig¨um¨us¸, B. Iyer, C. Li, and S. Mehrotra, “Executing sql over encrypted data in the database-service-provider model,” in Proc. ACM SIGMOD Int’l Conf. Management of data, June 2002.

9) L. Ferretti, M. Colajanni, and M. Marchetti, “Distributed, concurrent, and independent access to encrypted cloud databases,” IEEE Trans. Parallel and Distributed Systems, vol. 25, no. 2, Feb. 2014.

10) R. A. Popa, C. M. S. Redfield, N. Zeldovich, and H. Balakrishnan, “CryptDB: protecting confidentiality with encrypted query processing,” in Proc. 23rd ACM Symp. Operating Systems Principles, Oct. 2011.

11) C. Gentry, “Fully homomorphic encryption using ideal lattices,” in Proc. 41st ACM Symp. Theory of computing, May 2009.

12) A. Boldyreva, N. Chenette, and A. O’Neill, “Order-preserving encryption revisited: Improved security analysis and alternative solutions,” in Proc. Advances in Cryptology – CRYPTO 2011. Springer, Aug. 2011.

13) P. Paillier, “Public-key cryptosystems based on composite degree residuosity classes,” in Proc. Advances in Cryptology – EUROCRYPT99. Springer, May 1999.

14) D. Song, D. Wagner, and A. Perrig, “Practical techniques for searches on encrypted data,” in Proc. IEEE Symposium on Security and Privacy., May 2000.

15) L. Ferretti, F. Pierazzi, M. Colajanni, and M. Marchetti, “Security and confidentiality solutions for public cloud database services,” in Proc. Seventh Int’l Conf. Emerging Security Information, Systems and Technologies, Aug. 2013.

16) A. Greenberg, J. Hamilton, D. A. Maltz, and P. Patel, “The cost of a cloud: research problems in data center networks,” SIGCOMM Comput. Commun. Rev., vol. 39, no. 1, pp. 68–73, Jan. 2008.

17) L. Popa, S. Ratnasamy, G. Iannaccone, A. Krishnamurthy, and I. Stoica, “A Cost Comparison of Data Center Network Architectures,” in Proc. ACM Int’l Conf. Emerging Networking Experiments and Technologies, 2010.

18) R. N. Calheiros, R. Ranjan, A. Beloglazov, C. A. De Rose, and R.

Buyya, “Cloudsim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms,” Software: Practice and Experience, vol. 41, no. 1, pp. 23–50, 2011.

19) O. Goldreich, Foundations of Cryptography: Volume 2, Basic Applications. Cambridge university press, 2004.

20) Amazon Web Services, “Amazon Relational Database Service,” http://aws.amazon.com/rds, Mar. 2014.

21) EnterpriseDB, “Postgres Plus Cloud Database,” http: //enterprisedb.com/cloud-database, Mar. 2014.

22) Microsoft, “Windows Azure SQL Database,” http: //www.windowsazure.com/en-us/services/data-management, Mar. 2014.

23) Rackspace, “Rackspace Cloud Database,” http://www.rackspace. com/cloud/databases, Mar. 2014.

24) Amazon RDS Pricing, “Amazon Relational Database Pricing,” http://aws.amazon.com/rds/pricing, Mar. 2014.