

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Special Issue 2, March 2018

Secure and Efficient Cloud Data Deduplication

B.S.Dhanyashri¹, S.Kiruthiha², P. Shanmathy³, A.V. Kalpana⁴

U.G. Student, Department of Computer Engineering, Velammal Institute of Technology, Chennai,
Tamil Nadu, India^{1,2,3}

Assistant Professor, Department of Computer Engineering, Velammal Institute of Technology, Chennai,
Tamil Nadu, India⁴

ABSTRACT: Data deduplication enables data storage systems to find and remove duplication within data without compromising its availability. The goal of data deduplication is to store more data in less space by storing and maintaining files into a single copy, where the redundant copies of data are replaced by a reference to this copy. Remote cloud storage have become an indispensable part for various applications in nowadays network, which store a large amount of data and provide the partial data needed. With the rapid growing of cloud storage services, such as cloud storage, encryption becomes an important technique for protecting the confidentiality of data. To upgrade the security of deduplication and ensure the information classification by changing the anticipated message into an erratic message. "Data encryption provides an important guarantee for the security and privacy of clients' data; it limits the manners of the accessibility and availability of the encrypted data.

KEYWORDS: Cloud storage service, confidentiality, data encryption, security, privacy, accessibility

I. INTRODUCTION

Cloud has become a very important platform for information storage and processing. It modify primarily unlimited resources and delivers elastic services to end users. Cloud computing is to permit users to require enjoy all of those technologies, while not the necessity for deep experience with all of them. Since 2006, the development rate of worldwide data has outperformed the anticipated straight inclination and exponential information development has brought about information deduplication procedures turn out to be progressively prevalent in both industry and academia. Thus far, deduplication innovation has been ended up being to be remarkably viable for reinforcement stockpiling.

Information mining systems are utilized as a part of numerous exploration regions, including science, artificial intelligence, hereditary qualities and showcasing. While data mining techniques are a means to drive efficiencies and predict customer behavior, if used correctly, a business can set itself apart from its competition through the use of prescient analysis. Web mining, a sort of information mining utilized as a part of client relationship administration, incorporates data accumulated by customary information mining strategies and procedures over the web. Web mining aims to understand customer behavior and to evaluate how effective a particular website is. Other data mining techniques include network approaches based on multitask learning for classifying designs, guaranteeing parallel and versatile execution of information mining calculations, the mining of extensive databases, the treatment of social and complex information writes, and machine learning. Machine learning is a type of data mining tool that designs specific algorithms from which to learn and predict.

We display a propelled plan to help more grounded security by encoding the record. To enhance the security of deduplication and protect the data confidentiality. This solution aids in saving the storage space as well as minimize bandwidth requirements. The application helps in easy maintenance of data on the cloud platform. Sales and marketing

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Special Issue 2, March 2018

departments can mine customer data to improve lead conversion rates or to create one-to-one marketing campaigns. Data mining information on historical sales patterns and customer behaviors can be used to build prediction models for future sales, new products and services.

II .LITERATURE REVIEW

In 2016,Zhihua Xia, Member, IEEE, Xinhui Wang, Xingming Sun, Senior Member. It builds an exceptional tree-based file structure and propose a "Greedy Depth-first Search" calculation to give proficient multi-watchword positioned lookGreedy DFS algorithm is used to achieve higher search efficiency.Two search scheme is used to protect privacy.

In 2016,ZhirongShen, JiwuShu, and Wei Xue.In KSAC ,an primitive called HPE to enforce fine grained access control and perform multi field search query. File is retrieved whwn keyword matches query.

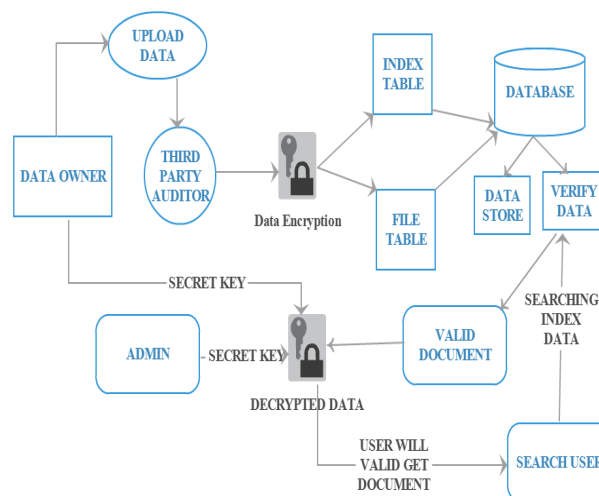
In 2017,Jianwei Yin, Yan Tang, Shuiguang Deng, Ying Li, and Albert Y. Zomaya, Fellow, IEEE.CTA amd DPE mechanism is used in dual-phase deduplication which runs on dynamic mode based on user’s input during runtime.

In 2017,Jiguo Li, Yao Wang, Yichen Zhang and Jinguang Han.ABEis a popular technology that is used to protect the security of usersdata.so, it is important to check the rightness of outsourced decoding to guarantee security for clients' information.

In 2016,Yucheng Zhang, Dan Feng, Member, IEEE, Hong Jiang, Fellow, IEEE, Wen Xia.This paper proposes another CDC calculation called the Asymmetric Extremum (AE) algorithm.AE has higher lumping throughput, littler piece estimate change than the current CDC calculations, and can discover appropriate piece limits in low-entropy strings

III. PROPOSED SYSTEM

It propose a scheme to de-duplicate encrypted data stored in cloud.It define the exact security guarantees required by such a data lineage mechanism toward identification of a guilty entity, and identify the simplifying non-repudation and honesty assumptions. Develop and analyze a novel “De-duplication “Technique on Encrypted cloud data, For that trusted data checker is used .Job of the data checker is to find given data is duplicate or if not then that data forwarded to key generator section for encryption as user specified algorithm. That data is separate into multi part and generate a signature key for each part, key based on part content.so, we can easy to identify data deduplication.



International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Special Issue 2, March 2018

A.DATAOWNER

Information proprietor can transfer data's, that information are part into multipart information at that point send to trusted information checker, employment of the information checker is to produce signature key from MD5 and compare with previous keys, if mismatch then that data send to Key generator Server, Job of the key generator are generate encryption key as user specified algorithm, finally encrypt then stores in Database. Data owner will upload the file based on the category in CSP(Cloud Service Provider).

B.DATA OWNER DATASET

This Module creates data owner dataset, this dataset only map owner with our uploaded data, it maintains common database to effectively find duplications. The server will use pairing computation over the whole database to realize the equality test. The equality test algorithm will finally output the correct result. The files will be uploading only once. If another data owner going to upload the same file in database means they will get the notification (the data is already uploaded in database). Data owner can save cost and time.

C.VERIFIER

In this modules, the third party auditor checks for the file integrity. If the file contains the same word as was in the file previously saved in the cloud then file will not store instead it shows error. TPA will filter the file. If the file has some updation with uniqueness then TPD will accept the file and encrypt the file and stored to the cloud. If the file is unique then that file will be directly stored in the cloud. Since the data items and the deduplication decision datasets are randomly generated, the storage time of each data item is usually not the same each time.

D.SHARED DATASET

Share Dataset is an light weight dataset that only contain mapping file metadata information. In this project, it maintain one common big data database instead of unique because it efficiently discovers duplication and memory association, if information proprietor share our information to customer that information not replicate rather portray name. Data deduplication enables data storage systems to find and remove duplication within data without compromising its availability.

E.SECURITY

To preserve the security in datasets, the file is encrypted and then the file is stored to the storage server. So if any authorized user request the file, the storage will initially declined their request or else if any user get the file, the files is encrypted and it is not in a readable format. It implements "Dynamic Encryption key Generation". It implies every single shared datum just view with information proprietor authorization, so we can dodge from obscure access. Social users are group members they can only view and share the data. If data user wants to show the data, they need to get permission to data owner then data owner will send Encryption key after they can view the data. If data owner does not provide the KEY mean user cannot view the file. Data encryption provides an important guarantee for the security and privacy of clients data, it limits the manners of the accessibility and availability of the encrypted data.

ALGORITHM:

AES (acronym of Advanced Encryption Standard) is a symmetric encryption algorithm. The algorithm was developed by two Belgian cryptographer Joan Daemen and Vincent Rijmen. AES was intended to be effective in both equipment and programming, and backings a piece length of 128 bits and key lengths of 128, 192, and 256 bits

AES ALGORITHM:

```
String name=res.getString(1);  
Byte[] na=name.getBytes();  
KeyGeneratorkeygenerator= KeyGenerator.getInstance("AES") SecretKeymyDesKey = keygenerator.generateKey()  
Cipher desCipher;  
desCipher = Cipher.getInstance("AES");
```

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Special Issue 2, March 2018

```
desCipher.init(Cipher.ENCRYPT_MODE,myDesKey);
byte[] na1= desCipher.doFinal(na);
```

IV. RESULT

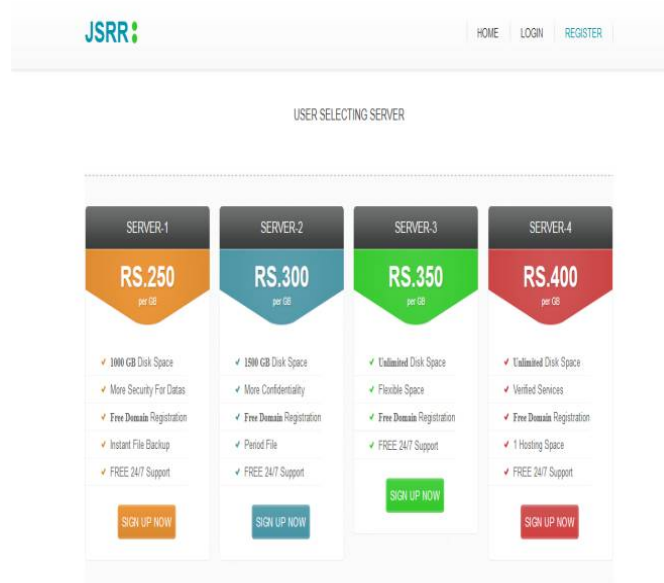


Fig 1. Web Page

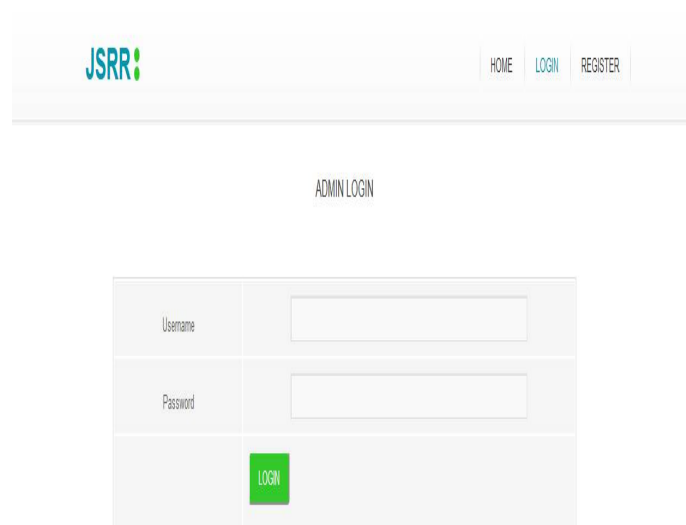


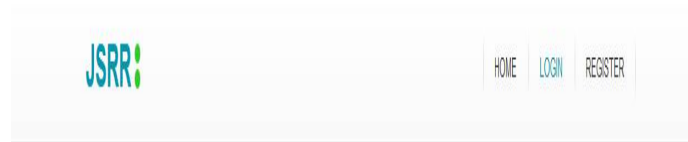
Fig 2. Admin Login

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Special Issue 2, March 2018



DATA OWNER LOGIN

E-Mail ID	<input type="text"/>
Password	<input type="password"/>
<input type="button" value="RESET ALL"/>	<input type="button" value="LOGIN"/>

Fig 3. Data Owner Login



DATA USER LOGIN

E-Mail ID	<input type="text"/>
Password	<input type="password"/>
<input type="button" value="RESET ALL"/>	<input type="button" value="LOGIN"/>

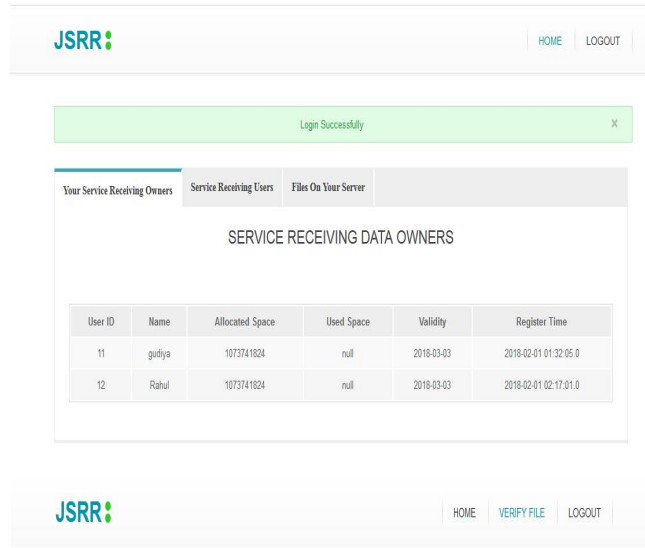
Fig 4. Data User Login

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Special Issue 2, March 2018



JSRR: HOME LOGOUT

Login Successfully X

Your Service Receiving Owners Service Receiving Users Files On Your Server

SERVICE RECEIVING DATA OWNERS

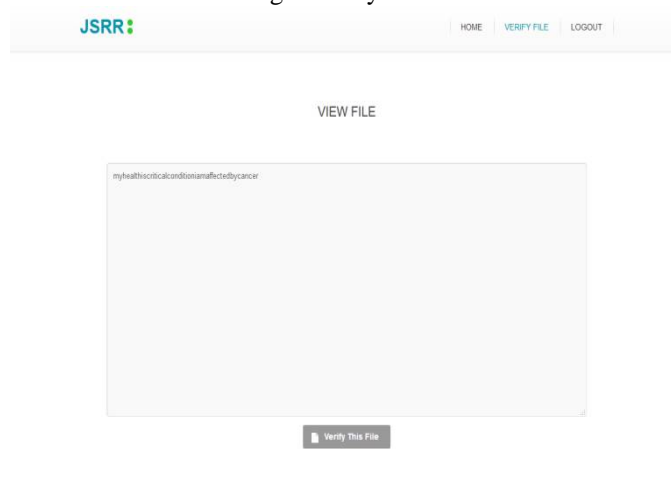
User ID	Name	Allocated Space	Used Space	Validity	Register Time
11	gudiya	1073741824	null	2018-03-03	2018-02-01 01:32:05.0
12	Rahul	1073741824	null	2018-03-03	2018-02-01 02:17:01.0

JSRR: HOME VERIFY FILE LOGOUT

WANT TO VERIFY FILES

Data ID	Document Name	Size	Indexes	Added Time	View Data	Verify Data	Reject Data
10	LICENSE	1088	10_LICENSE	2017-02-04 12:21:13.0	View Data	Verify	Reject
17	New Text Document.txt	32	17_New Text Document.txt	2017-03-22 07:59:30.0	View Data	Verify	Reject

Fig 5. Verify Files



JSRR: HOME VERIFY FILE LOGOUT

VIEW FILE

myhealthcriticalconditionisaffectedbycancer

Verify This File

Fig 6. View File

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Special Issue 2, March 2018

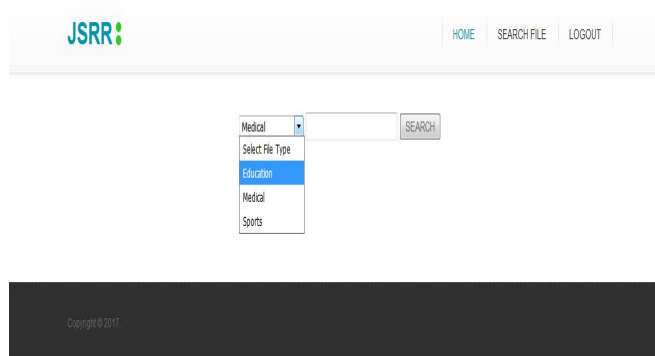


Fig 7.Search File

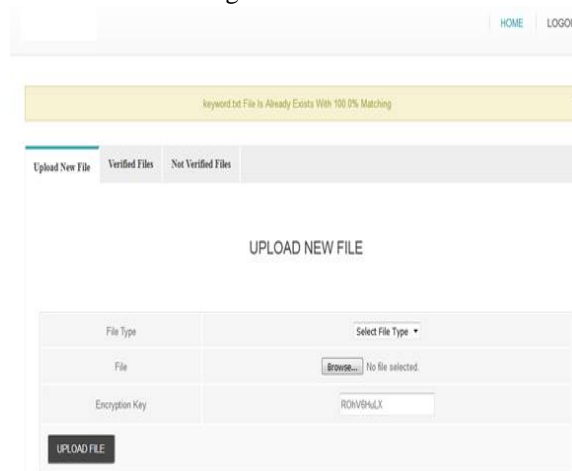


Fig 8.Deduplication Find

V. CONCLUSION

Cloud computing is viewed as the next generation architecture of IT companies. Previous work on privacy preserving quantitative association rule mining using discrete wavelet transform has been set in an environment where the collection of data is not transient and not subject to change. As promising as it is, cloud computing also brings forth many new security issues when users outsource sensitive data to cloud servers. To keep delicate clients' information private against untrusted servers, existing arrangements for the most part apply cryptographic techniques. With data encryption, the same file will become different from each other, thus deduplication which is widely adopted by cloud storage service providers meets some challenges. Current strategy to tackle the issue is to make utilization of some data registered from the mutual record to accomplish deduplication of encoded information, say unit encryption. But this piece of information which is computable from the file via a deterministic public algorithm is not really meant to be secret. To this end, we propose a plan to address the deduplication of encoded information productively and safely with

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Special Issue 2, March 2018

the assistance of guaranteeing the responsibility for shared document, scrambling data using keys at user's will and realizing the anonymous store through the digital credential.

REFERENCES

- [1] D.Song, D.Wagner, and A.Perrig, "Practical techniques for search on encrypted data," in Proc. of IEEE S&P, 2000, pp. 44–55.
- [2] Y.-C.Chang and M.Mitzenmacher, "Privacy preserving keyword search on remote encrypted data," in Proc. of ACNS, 2005, pp. 391–421.
- [3] C.Wang, N.Cao, K.Ren, and W.Lou, "Enabling secure and efficient ranked keyword search over outsourced cloud data," IEEE Trans. Parallel Distrib. Syst., vol. 23, no. 8, pp. 1467–1479, 2012.
- [4] P.Golle, J.Staddon, and B.R.Waters, "Secure conjunctive keyword search over encrypted data," in Proc. of ACNS, 2004, pp. 31–45.
- [5] D.Boneh and B.Waters, "Conjunctive, subset, and range queries on encrypted data," in Proc. of TCC, 2007, pp. 535–554.
- [6] Y.Hwang and P.Lee, "Public key encryption with conjunctive keyword search and its extension to a multi-user system," in Proc. of Pairing, 2007, pp. 2–22.
- [7] E.Shen, E.Shi, and B.Waters, "Predicate privacy in encryption systems," in Proc. of TCC, 2009, pp. 457–473.
- [8] W.K.Wong, D.W.Cheung, B.Kao, and N. Mamoulis, "Secure knn computation on encrypted databases," in Proc. of ACM SIGMOD, 2009, pp. 139–152.
- [9] N.Cao, C.Wang, M.Li, K.Ren, and W.Lou, "Privacy-preserving multi-keyword ranked search over encrypted cloud data," in Proc. of INFOCOM, 2011, pp. 829–837.
- [10] J.S.Vaidya and C.Clifton, "Privacy-preserving multi-keyword ranked search over encrypted cloud data," IEEE Trans. Parallel Distrib. Syst., vol. 25, no. 1, pp. 222–233, 2014.
- [11] S. Kamara, C. Papamanthou, and T. Roeder, "Dynamic searchable symmetric encryption," in Proc. of the ACM Conference on Computer and Communications Security, NC, USA, Oct. 2012, pp. 965–976.
- [12] S. Kamara and C. Papamanthou, "Parallel and dynamic searchable symmetric encryption," in Proc. of Financial Cryptography, Okinawa, Japan, Apr. 2013, pp. 258–274.
- [13] M. Naveed, M. Prabhakaran, and C. Gunter, "Dynamic searchable encryption via blind storage," in Proc. of IEEE Symposium on Security and Privacy, CA, USA, May 2014, pp. 639–654.
- [14] R. Agrawal, J. Kiernan, R. Srikant, and Y. Xu, "Order preserving encryption for numeric data," in Proc. of ACM SIGMOD, Paris, France, Jun. 2004, pp. 563–574.
- [15] H. Hacigumus, B. Iyer, C. Li, and S. Mehrotra, "Executing sql over encrypted data in the database-service-provider model," in Proc. of ACM SIGMOD, Madison, Wisconsin, Jun. 2002, pp. 216–227.
- [16] H. Kadhemi, T. Amagasa, and H. Kitagawa, "A secure and efficient order preserving encryption scheme for relational databases," in Proc. of the International Conference on Knowledge Management and Information Sharing, Valencia, Spain, Oct. 2010, pp. 25–35.
- [17] R. A. Popa, C. M. S. Redfield, N. Zeldovich, and H. Balakrishnan, "Cryptdb: Protecting confidentiality with encrypted query processing," in Proc. of ACM Symposium on Operating Systems Principles, Cascais, Portugal, Oct. 2011, pp. 85–100.
- [18] R. A. Popa, F. Li, and N. Zeldovich, "An ideal-security protocol for order-preserving encoding," in Proc. of IEEE Symposium on Security and Privacy, CA, USA, May 2013, pp. 463–477.
- [19] X. Chen, J. Li, J. Ma, Q. Tang, and W. Lou, "New algorithms for secure outsourcing of modular exponentiations," IEEE Transactions on Parallel and Distributed Systems, vol. 25(9), pp. 2386–2396, Jul. 2014.
- [20] X. Chen, J. Li, J. Weng, J. Ma, and W. Lou, "Verifiable computation over large database with incremental updates," in ESORICS 2014, ser. Computer Science. Springer-Verlag, 2014, vol. 8712 of LNCS, pp. 148–162.