May- June 2020
ISSN: 0193-4120 Page No. 6880 - 6884

Secure Auditing and Deduplicating in Cloud Computing using MD5 Algorithm

Dr Sujaritha[1].M, Sibimanigandan.S[2], Sachin Mohammed Rafic[3], Sibi Athithya.C.A[4], Sujith.A[5]
[1]Professor, [2,3,4,5]Student, Department of Computer Science Engineering,
Sri Krishna College of Engineering and Technology, Coimbatore, India.

Article Info
Volume 83
Page Number: 6880 - 6884
Publication Issue:
May-June 2020

Article History
Article Received: *19 November 2019*
Revised: *27 January 2020*
Accepted: *24 February 2020*
Publication: *18 May 2020*

Abstract:
This paper depicts about bringing a change in algorithm for the IEE paper "Secure auditing and deduplicating data in cloud computing". This paper intends to reduce duplicated data. It uses MD5 algorithm for generating hash function during transaction to cloud. The MD5 algorithm splits the input data into 512kbs and then transmits, thus making the data transition simple and unbreachable. Even if the third party tries to access the data it would get only the part of the data since MD5 transmits data only with the size of 512kbs. The data which is been sent to the cloud is encrypted and then sent, thus making the data secure. Similarly during accessing those data the encrypted data is been decrypted and then accessed. These data which are been stored in the cloud can be accessed from anywhere around the world with the help of ONEDRIVE app. This paper uses ONEDRIVE by Microsoft as its cloud source. ONEDRIVE is available for both Ios and Android devices. By using this algorithm the data are saved and protected much better than the prior method. This algorithm ensures that the data are not been duplicated and it also provides maximum security. The working of this paper is tested and result is obtained successfully.

Keywords*: Cloud storage, data de-duplicating, secure auditing, and MD5 algorithm.*

Introduction:

The development of cloud computing during the past decade has been enormous. Most of the companies both MNC's and small scale companies have been switching over the cloud computing for storing data and accessing data. Before cloud, there would be lot of storage device for a single company and it would be hard to retrieve those data. Cloud provides us with a option i.e. it allows its user to access the data and store data wherever it is possible. With the arrival of

Cloud into the world it has been simpler for storing and accessing data. Even though cloud has provided simpler storage and accessing methods it has quite thin firewall. Cloud sure does shows vulnerability when it comes to security and data also gets duplicated once in a while. Duplication of data results in large accumulation of repeated data and thus occupying more space than it actually requires. Thus making companies run out space. This paper proposes a solution i.e.MD5 algorithm. This algorithm ensures that the data are not been duplicated and also provides security to data without being mugged by third parties. This algorithm encrypts the data before transition and then sends it to the cloud. MD5 splits the data into smaller parts (512kbs) and then transmits, Thus making the data secured. Hash functions are generated during the encryption and decryption process. This is done to check whether the data is intact. This paper is about a windows application which uses java programming, One drive cloud and MD5 algorithm to make the data stored in the cloud secured and avoid duplication. One drive is a free cloud service that is been provide by Microsoft. With this we can access the data wherever and whenever it is possible.

II.METHODOLOGY:

Mainly the project is a concept to avoid duplication of data and providing security to it. The updates that is been done to the data can be viewed using One

May- June 2020
ISSN: 0193-4120 Page No. 6880 - 6884

Drive app. The project is an windows based application which is programmed using java. The windows application runs on java with swing components, Swing components provides better interface than applets. Initially the data that is said to be transferred is encrypted using MD5 algorithm. Later this encrypted data is uploaded to third party administrator. Md5 algorithm provides security to the data as it splits each data into the size of 512kbs. During the process of encryption it generates a hash function which is unique for each data. By this process it would be hard for any third party to access data without proper authorization. MD5 transfers data of size 512kbs during encryption and similarly it does the same during decryption. Only registered users are made to access data in this project. During the process of decryption it generates a hash function which would be similar to the hash function which was generated initially. The data would get decrypted only when the hash functions are similar. If the hash functions do not match then the data won't be encrypted as it might be mugged by the third party. Thus this provides maximum security to the data. Data does not get duplicated since if there is a file already in the cloud with same name then the algorithm does not allows it from entering the cloud . It would reject the entry of duplicated data.

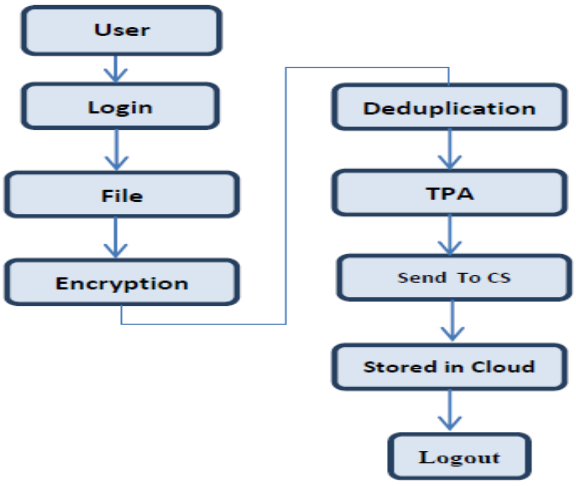
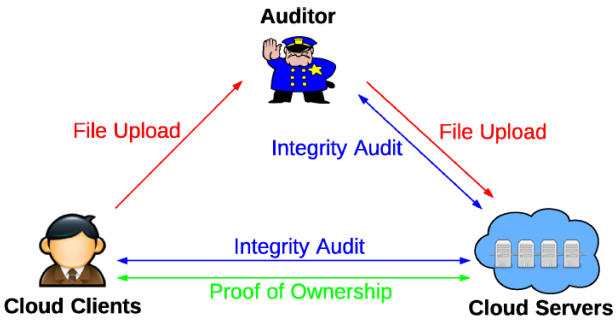


Figure.1. Block Diagram

III SYSTEM ARCHITECTURE:



As shown in the diagram the system is been built. The client who wishes to access the cloud is subjected to a auditor who intend allows access only for users with proper registration thus providing high level security on data. By doing so we can ensure that these data are secured and they are of high integrity. The request that is made by the user is first encrypted before transmission by doing so it can be secured. Similarly on the receivers end the encrypted request is decrypted into readable form thus providing the client with undisputed messages.

2. MYSQL:

This project sure did need a database for storing and accessing data thus we have use MYSQL. This provides us with the following features-

- **Relational Database Management System (RDBMS):**
  MySQL is a social database the board framework.
- **Easy to utilize:**
  MySQL is anything but difficult to utilize. You need to get just the fundamental information on SQL. . You can create and interface with MySQL with just a couple of clear SQL articulations.
- **Security provided:**
  MySQL include a solid data security layer that shields unstable data from interlopers. Passwords are converted into code in MySQL.
- **Client/ Server Architecture:**
  MySQL follows a customer/server engineering. There is a database server (MySQL) and discretionarily various clients (application programs), which talk with the

May- June 2020
ISSN: 0193-4120 Page No. 6880 - 6884

server i.e., data are been checked, changes in data are noted and saved, etc.

- **It is adaptable:** MySQL can deal with practically any measure of information, up to as much as 50 million lines or more. The default document size breaking point is around 4 GB. In any case, you can build this number to a hypothetical restriction of 8 TB of information.
- **Compatibility**: MySQL is good to run on many working frameworks, as Novell NetWare, Windows Linux, various groupings of UNIX, (for example, Sun Solaris, AIX, and DEC UNIX), operating system/2, FreeBSD, and others. MySQL additionally gives an office that the customers can run on a similar PC as the server or on another PC (correspondence by means of a nearby system or the Web).
- **Provides roll back feature:** MySQL permits exchanges to be moved back, submit and crash recuperation.
- **Includes:** MySQL is faster, dynamically strong and more affordable because of its novel amassing engine designing.

V.SOFTWARE DEVELOPMENT:

A Windows based application is needed to be built to make the data secured and remove duplication. This project uses MD5 algorithm which provides the data with security and helps to remove duplicated data. Initially a login page is been built using java swing as it provides better interface. This allows only registered users to access data and thus minimizes the chance of data theft and it then creates keys. Mystery key introduction can occur right now. An inspecting convention with key presentation versatility is made by five calculations, for example, Sys Arrangement, Key Update, Auth Gen, Evidence Gen, and Confirmation Check. Current security model considers the idea of the forward security and information ownership property. The cloud is permitted to get the customer's mystery key for distributed storage examining. Once the cloud gets access to the key it then allows the user to upload data. The customer produces records and transfers these documents alongside relating authenticators to the cloud. The cloud stores these records for the customer and gives download administration if the customer requires. Dynamic information activities for review administrations are likewise gone to so as to make examining increasingly adaptable. The cloud safeguards these documents for the customer and gives download administration if the customer requires. Each record is further increasingly isolated into various squares. For the effortlessness of portrayal, the customer can intermittently review whether his documents in cloud are right. The customer will refresh his mystery keys for distributed storage reviewing toward the finish of each timeframe, yet the open key is constantly unaltered. The cloud is permitted to get the customer's mystery key for distributed storage examining in one certain time span. It implies the mystery key presentation can occur right now evaluating convention accomplishes key-introduction versatility while fulfilling current proficiency necessities. It is authorized in the definition and the security model of evaluating convention with key-presentation versatility, and has given the pragmatic arrangement. The security confirmation and the asymptotic introduction evaluation portrayed that the convention is secure and proficient. In performance evaluation, implement the graph for our proposed work which compares the existing algorithm with proposed work algorithm.

VI.RESULT AND DISCUSION:

Images of output:

May- June 2020
ISSN: 0193-4120 Page No. 6880 - 6884

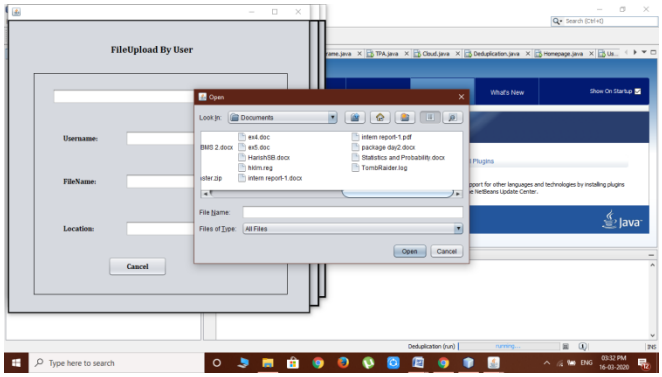


Fig.3 User uploading file

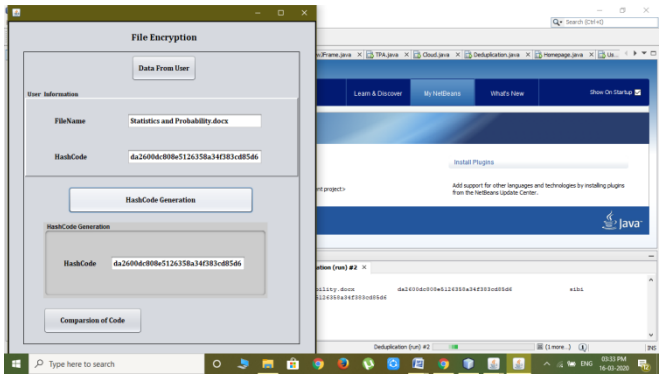


Fig.4 Generating Hash Function

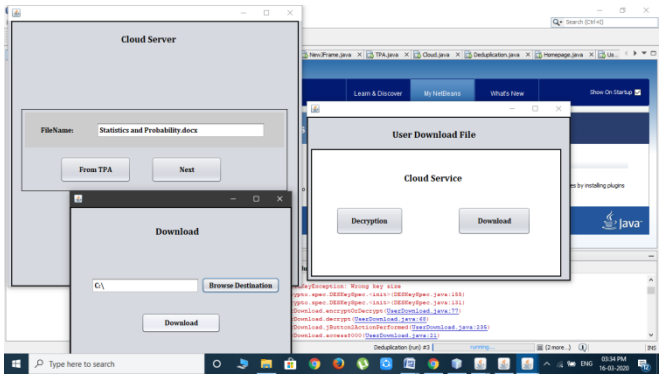


Fig.5 Downloading File after decryption

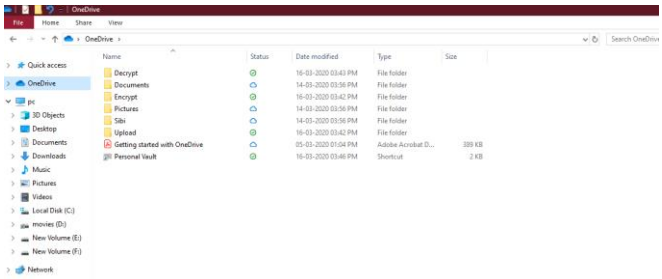


Fig.6 Saved file in the cloud (ONEDRIVE)

CONCLUSION

The paper focuses on providing the data with security and minimizing the duplication. This paper uses MD5 algorithm and it helps in minimizing the duplication of data and also provides security to data. The application build is a windows based application, which includes user registration. Thus making the data secure as the registered users are the only ones who can access data. The data are been encrypted during the process of transmitting and they are decrypted during the process of decryption. Same data files are not accepted as it would become duplicates of the original data and occupy large space. If same files are uploaded then the application does not take it into account. Cloud computing provides lot of options but security is not one. The security it provides is very thin and it is quite simply breachable. Hash functions are been generated during the process to provide maximum security and to avoid duplication of data. Thus this paper provides us with secured data and minimizes duplicated data with the help of MD5 algorithm.

REFERENCES

1. J. R. Douceur, A. Adya, W. J. Bolosky, D. Simon,and M. Theimer,“Reclaiming space from duplicate files in a server less distributed file system.” in ICDCS,2002, pp. 617–624.
2. M. W. Storer, K. Greenan, D. D. E. Long, and E. L.Miller, “Secure data deduplication,” in Proc. of StorageSS, 2008.
3. P. Anderson and L. Zhang, “Fast and securelaptop backups with encrypted de-duplication,” in Proc. of USENIX LISA, 2010
4. M. Bellare, S. Keelveedhi, and T. Ristenpart,“Dupless: Server aided encryption for deduplicated storage,” in USENIX Security Symposium, 2013.
5. Science, IEEE, 1997.Jin Li, Yan Kit Li, Xiaofeng Chen, Patrick P. C. Lee, Wenjing Lou, “A Hybrid Cloud Approach for Secure Authorized Deduplication”, IEEE Transactions on Parallel and Distributed Systems, Volume: PP, Issue:99, Date of Publication :18.April.2014
6. A. Adya, W. J. Bolosky, M. Castro, R. Chaiken, G. Cermak, J. R. Douceur, J. Howell, J. R. Lorch, M. Theimer, and R. Wattenhofer. FARSITE: Federated, available, and reliable storage for an incompletely trusted environment. In Proceedings

May- June 2020
ISSN: 0193-4120 Page No. 6880 - 6884

of the 5th Symposium on Operating Systems Design and Implementation (OSDI), Boston, MA, Dec.2002.USENIX.

7. R. Anderson and E. Biham, "Two Practical and Provably Secure Block Ciphers: BEAR and LION", 3rd International Workshop on Fast Software Encryption, 1996, pp. 113-120.
8. P. Golle, S. Jarecki, and I. Mironov.Cryptographic primitives enforcing communication and storage complexity.In "Financial Cryptography "02", volume 2357 of LNCS, pages 120–135. Springer, 2003.
9. A. Juels and B. S. Kaliski, Jr. Pors: proofs of retrievability for large files. In ACM CCS "07, pages 584– 597. ACM, 2007
10. H. Shacham and B. Waters. Compact proofs of retrievability. In ASIACRYPT "08, pages 90–107. SpringerVerlag, 2008.
11. A.D. Santis and B. Masucci, ""Multiple Ramp Schemes,"" IEEE Trans. Inf. Theory, vol. 45, no. 5, pp. 1720-1728, July 1999.
12. G.R. Blakley and C. Meadows, ""Security of Ramp Schemes,"" in Proc. Adv. CRYPTO, vol. 196, Lecture Notes in Computer Science,G.R. Blakley and D. Chaum, Eds., 1985, pp. 242-268.
13. M.O. Rabin, ""Efficient Dispersal of Information for Security, Load Balancing, Fault Tolerance,"" J. ACM, vol. 36, no. 2, pp. 335- 348, Apr. 1989.
14. A. Shamir, ""How to Share a Secret,"" Commun. ACM, vol. 22, no. 11, pp. 612-613, 1979.
15. J. Gantz and D. Reinsel, the Digital Universe in 2020: Big Data, Bigger Digital Shadows, Biggest Growth in the Far East, Dec. 2012. [Online]. Available: http://www.emc.com/collateral/analystreports/idc-thedigitaluniverse-in-2020.pdf.