iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 6 **Issue:** IV **Month of publication:** April 2018

DOI: http://doi.org/10.22214/ijraset.2018.4356

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue IV, April 2018- Available at www.ijraset.com

A Secure Access Policies Based Data Deduplication System

Hari Shankar Prasad. R[1], Mr. S.P. Maniraj[2]
[1]*Department Of CSE, SRMIST, Chennai-89*
[2]*Asst.pro, Department Of CSE, SRMIST, Chennai-89*

Abstract: ***This paper presents a survey on an attribute-based storage system with secure deduplication in a hybrid cloud setting with higher confidentiality and reliability. Deduplication is a technique which is widely employed to backup data and minimizes network and storage overhead by detecting and eliminating redundancy among data. It eliminates duplicate copies of identical data in order to save storage space and network bandwidth. Private cloud is responsible for duplicate detection and a public cloud manages the storage. Instead of keeping multiple data copies with the same content, in this system eliminates redundant data by keeping only one physical copy and referring other redundant data to that copy. Each such copy can be defined based on user access policies, the user will upload the file with access policies and then file type question with answer. Then same file with different access policies to set the particular file to replace the reference. Applying deduplication to user data to save maintenance cost in HDFS storage system.***

Index Terms: ***Cloud Computing, Encryption Techniques, Data security, Authentication, Security in Cloud.***

I. INTRODUCTION

Cloud Computing is a combination of IT services provided by many service providers. The term cloud was originated from the internet and is also a platform that gives people the opportunity for sharing resources, services and information globally. In general, cloud computing has diverse definitions obtained by several important organizations. With infinite storage space provide by cloud service provider users tend to use as much space as they can and vendors constantly look for techniques aimed to minimize redundant data and maximize space savings. Users will access information according to their needs and most users access same information again and again, the cost of computation, application hosting, content storage and delivery is reduced significantly.A technique has been introduced for reducing the amount of storage space an organization needs to save its data which is the data deduplication system. It helps in eliminating the storage systems contain duplicate copies of many pieces of data. For example, the same file may be saved in several different places by different users, or two or more files that aren't identical may still include much of the same data. Along with low ownership costs and flexibility, users require the protection of their data and confidentiality guarantees through encryption. To make data management scalable deduplication we are use Encryption for secure deduplication services. Both the deduplication system and encryption system are two conflicting technologies present. The cipher text and the user's private key depend on attributes. If the features of a user key counterpart to those of the cipher text, then decryption is permitted. ABE utilizes four algorithms: encryption, decryption, setup and key generation.

The DES algorithm ensures data security in cloud. The security design architecture of the system is planned by using DES cipher block chaining, which eradicates the hackers. The data which is sent, being interrupted and replaced has no danger. The system with encryption is adequately secure but for that, the level of encryption has to be amplified, as computing power upsurges. To secure the results, a symmetric key is used to encrypt the communication system between the modules. The cloud data security must be utilized to analyse the data security requirements, the data security risk, the data security process through encryption and disposition of security functions.

This encryption algorithm is used to address the security and privacy issues in cloud storage to shield the data from illicit access.

II. RELATED WORK

There are several proposed models in the recent years based on Secure Deduplication in Cloud Storage. Our system aims to enable the deduplication effectively and avoid redundant datas. The author perfoms the deduplication to spare space by just putting away one duplicate of every document or file transferred. Should customers routinely scramble their documents, be that as it may, funds are lost. Message-bolted encryption (the most unmistakable appearance of which is concurrent encryption) certify this strain. In any case it is intrinsically subject to savage power assaults that can recoup records falling into a known set. In Dupless, customers encode under message-based keys acquired from a key-server by means of an absent PRF convention. We demonstrate that

encryption for deduplicated stockpiling can accomplish execution and space reserve funds near that of consuming the stockpiling administration with plaintext information. Different types or individual data now store extensive amounts of individual and corporate information on tablets or home PCs. By doing this type of work it is helpless against burglary or equipment disappointment. Ordinary ideal arrangements are not appropriate to this environment, and reinforcement administrations are every now and again deficient. This system depicts a calculation which exploits the information which is basic between clients to build the pace of reinforcements, and decrease the capacity necessities. Deduplication is a system for taking out copy duplicates of information, and has been broadly utilized as a part of distributed storage to decrease storage space and transfer data transfer capacity. Promising as it perhaps, an emerging test is to perform secure deduplication in distributed storage. This system makes the first endeavor to formally notify the issue of accomplishing effective and dependable key administration in secure deduplication. Firstly we introduce a pattern approach in which every client holds an autonomous expert key for scrambling the aim keys and outsourcing them to the cloud. On the second way, such a standard key administration plan produces a tremendous number of keys with the expanding number of obliges clients and clients to dedicatedly secure the expert keys. To this end, we propose Dekey , another development in which clients don't have to deal with any keys all alone however rather safely circulate or transfer the united key shares over different servers. Security examination exhibits that Dekey is secure as far as the definitions determined in proposed security model. As a proof of idea, we actualize Dekey utilizing the Ramp mystery sharing plan and show that Dekey brings about restricted overhead in reasonable situations. Distributed storage frameworks are turning out to be progressively prominent. A promising innovation that holds their expense down is de-duplication, which stores just a solitary duplicate of rehashing information. Customer side deduplication endeavors to recognize deduplication opportunities as of now at the customer and save the transmission capacity of transferring duplicates of the existing documents or files to the server. Another idea namely call private information deduplication convention, a deduplication system for private information stockpiling is presented and formalized. Naturally, a private information deduplication convention allow a customer who holds a private information demonstrates to a server who have a synopsis string of the information that he/she is the proprietor of that information without uncovering additional data to the server. Our idea can be seen as a supplement of the cutting edge open information deduplication conventions of Halevy al.

A. Cloud based Storage Drive Forensics

As the data are increasing, the storage become major issue for the people. It perform forensics of Google Drive via di1fferent technique such as using client software, Google Drive access via browser, Memory Analysis, Network Analysis and other techniques.

B. Message-Locked Encryption for Lock-Dependent Messages

It helps in avoiding duplication in storage systems and Message-Locked Encryption (MLE) schemes which sub assumes convergent encryption and its variants. The schemes do not rely on permanent secret keys, but rather encrypt messages using keys de- rived from the messages themselves. It helps avoiding the use of ciphertext components derived deterministically from the messages and it deterministic ciphertext component that enables more

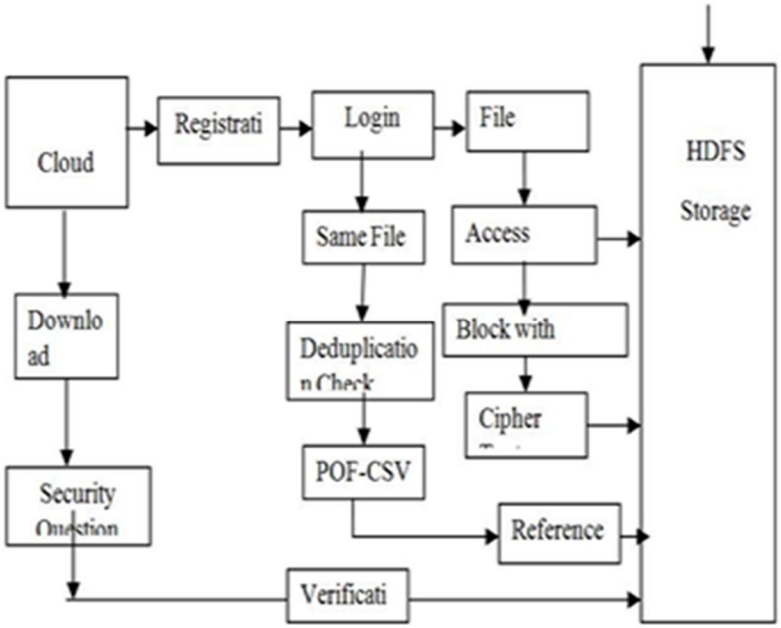
C. Message-Locked Encryption and Secure Deduplication

It secures deduplication, a goal currently targeted by numerous cloud-storage providers. It provide ROM security analyses of a natural family of MLE schemes that includes deployed schemes. The challenge is standard model solutions, and we make connections with deterministic encryption, hash functions secure on correlated inputs and the sample-then-extract paradigm to deliver schemes under different assumptions and for different classes of message sources.

III. PROPOSED MODELLING

The first module is the User registration Module in which the User has to upload his/her files in a cloud server and should register first before uploading. Then only he/she can be capable of doing it. For that he requires to fill the details in the registration form. The user details are maintained in a database. In the next module the user uploads the file with access policies ,in this module User will chooses the file and uploads to Storage where the HDFS storage system .In the system will generate a signature in particular file and then split into multiple block. Each block will be generate signature with key. The MD5 message-digest algorithm is cryptographic hash function which is typically expressed in text format as 32 digit hex value. After that generate convergent keys for each block splitting to store CSV file like filename, file path, blocks, username, password and block keys. The RSA algorithm helps in encypting the blocks is asymmetric cryptography algorithm. As the name describes that the Public Key is given to everyone and Private Key is kept private. Here the plain text is encryption to cipher text and stored in slave system. Blocks are stored in

Distributed HDFS Storage Providers. After upload the file to set the access policies with set security question. The next module is the Detection Deduplication method, File-level data deduplication compares a file to be backed up or archived with copies that are already stored and it is done by checking the attributes against an index. If the file is unique, it is been stored .And the index is updated; if only a pointer to the existing file is stored. The result is that only one instance of the file is saved, and subsequent copies are replaced with a reference that points to the original file. Another one signature match checking looks within a file and saves unique Iterations of each block. All the blocks are broken into compressed with the same fixed length. Each compressed data is processed using a hash algorithm such as MD5 or SHA-1. Download User File is the final model user request for downloading their own document which they have uploaded in HDFS storage. In this download request will analysis the user attribute once it will matched then ask the security questions for particular file. After complete the process needs proper ownership verification. The original content is decrypted by requesting the Distributed HDFS storage where HDFS storage request key management slave for keys to decrypt and finally the original content is received by the user.



IV. RESULTS AND DISCUSSIONS

We have proposed the system which helps in improving the reliability of data. It also helps in achieving the confidentiality of the user's data. The Unique feature of the proposal is that data integrity, as well as tag consistency, can be achieved. The secured system and data owner can decide whether the user can access the system or not.

1) *KeyGen(F):* The key generation algorithm gives a input as file content F and generate outputs as the convergent key ckF for F.
2) *Encrypt (ckF;F):* The encryption algorithm gives input as the convergent key ckF with file content F and generate the ciphertext ctF as output.
3) *Decrypt (ckF; ctF):* The decryption algorithm gives input, the convergent key ckF with ciphertext ctF and generate the plain file F as output
4) *Data Encryption Standard (DES):* DES is a *block cipher*--meaning it operates on plaintext blocks of a given size (64-bits) and returns ciphertext blocks of the same size.
5) *Example:* Let M be the plain text message M = 0123456789ABCDEF, where M is in hexadecimal (base 16) format. Rewriting M in binary format, we get the 64-bit block of text:

M = 0000 0001 0010 0011 0100 0101 0110 0111 1000 1001 1010 1011 1100 1101 1110 1111
L = 0000 0001 0010 0011 0100 0101 0110 0111
R = 1000 1001 1010 1011 1100 1101 1110 1111
The first bit of M is "0". The last bit is "1". We read from left to right.
DES operates on the 64-bit blocks using *key* sizes of 56- bits. The keys are actually stored as being 64 bits long, but every 8th bit in the key is not used (i.e. bits numbered 8, 16, 24, 32, 40, 48, 56, and 64). However, we will nevertheless number the bits from 1 to

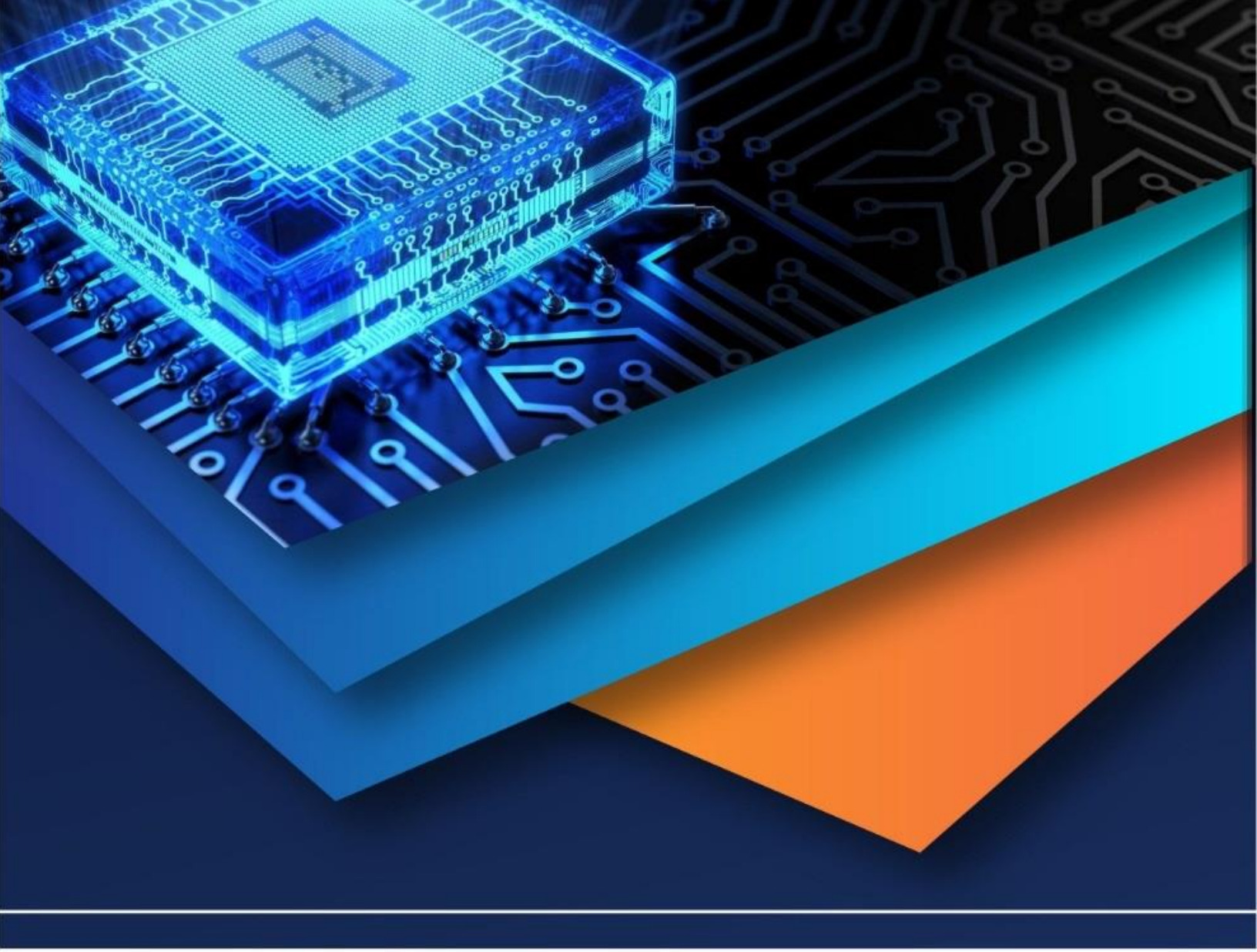
64, going left to right, in the following calculations. But, as you will see, the eight bits just mentioned get eliminated when we create subkeys.

V. CONCLUSION

To enable de-duplication in convergent keys and distribute the convergent keys across multiple Key Management slave machines the attributes And chunks in the various cloud storage providers and downloaded securely by receiving keys from Key Management slaves and Chunks from the Distributed HDFS with improved reliability. The deduplication and distributed storage of the data across HDFS is achieved .And then using two way cloud in our storage system is built under a hybrid cloud architecture, where a private cloud manipulates the computation and a public cloud manages the storage.. If so, whenever it is necessary, it regenerates the cipher text into a cipher text of the same plaintext over an access policy which is the union set of both access policies. Like public cloud and private cloud. We have shown the concept of deduplication effectively and security is achieved by means of Proof of Ownership of the file. That is attribute-based storage system cipher text-policy attribute-based encryption (CP-ABE) and supports secure deduplication. User privacy is enhanced by access requests to privately inform the cloud server about the users access desires .Forward security I realized by the session identifiers to prevent the session correlation. It is possibly applied for enhanced privacy preservation in cloud applications.

REFERENCES

[1] D. Quick and K. R. Choo, “Google drive: Forensic analysis of data remnants,” J. Network and Computer Applications, vol. 40, pp. 179–193, 2014.
[2] A. Sahai and B. Waters, “Fuzzy identity-based encryption,” in Advances in Cryptology - EUROCRYPT 2005, 24th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Aarhus, Denmark, May 22-26, 2005, Proceedings, ser. Lecture Notes in Computer Science, vol. 3494. Springer, 2005, pp. 457–473.
[3] M. Bellare, S. Keelveedhi, and T. Ristenpart, “Message-locked encryption and secure deduplication,” in Advances in Cryptology EUROCRYPT 2013, 32nd Annual International Conference on the Theory and Applications of Cryptographic Techniques, Athens, Greece, May 26-30, 2013. Proceedings, ser. Lecture Notes in Computer Science, vol. 7881. Springer, 2013, pp. 296–312.
[4] M. Abadi, D. Boneh, I. Mironov, A. Raghunathan, and G. Segev, “Message-locked encryption for lock-dependent messages,” in Advances in Cryptology - CRYPTO 2013 - 33rd Annual Cryptology Conference, Santa Barbara, CA, USA, August 18-22, 2013. Proceedings,Part I, ser. Lecture Notes in Computer Science, vol. 8042. Springer,2013, pp. 374–391.
[5] S. Keelveedhi, M. Bellare, and T. Ristenpart, “Dupless: Serve raided encryption for deduplicated storage,” in Proceedings of the 22th USENIX Security Symposium, Washington, DC, USA, August 14-16, 2013. USENIX Association, 2013, pp. 179–194.
[6] .M. Bellare and S. Keelveedhi, “Interactive message-locked encryption and secure deduplication,” in Public-Key Cryptography – PKC 2015 - 18th IACR International Conference on Practice and Theory in Public-Key Cryptography, Gaithersburg, MD, USA, March 30 – April 1, 2015, Proceedings, ser. Lecture Notes in Computer Science, vol. 9020. Springer, 2015, pp. 516–538.
[7] S. Bugiel, S. N¨ urnberger, A. Sadeghi, and T. Schneider, “Twin clouds: Secure cloud computing with low latency - (full version),”in Communications and Multimedia Security, 12th IFIP TC 6 / TC 11 International Conference, CMS 2011, Ghent, Belgium, October 19-21,2011. Proceedings, ser. Lecture Notes in Computer Science, vol.7025. Springer, 2011, pp. 32–44.
[8] V. Goyal, O. Pandey, A. Sahai, and B. Waters, “Attribute-based encryption for fine-grained access control of encrypted data,” in Proceedings of the 13th ACM Conference on Computer and Communications Security, CCS 2006, Alexandria, VA, USA, Ioctober 30 November 3, 2006, ser. Lecture Notes in Computer Science, vol.5126. Springer, 2006, pp. 89–98.
[9] J. Bethencourt, A. Sahai, and B. Waters, “Ciphertext-policy attribute-based encryption,” in 2007 IEEE Symposium on Security and Privacy (S&P 2007), 20-23 May 2007, Oakland, California, USA. IEEE Computer Society, 2007, pp. 321–334.
[10] M. W. Storer, K. M. Greenan, D. D. E. Long, and E. L. Miller, “Secure data deduplication,” in Proceedings of the 2008 ACM Workshop On Storage Security And Survivability, StorageSS 2008, Alexandria VA, USA, October 31, 2008. ACM, 2008, pp. 1–10.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY