iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 7 **Issue:** III **Month of publication:** March 2019

DOI: http://doi.org/10.22214/ijraset.2019.3080

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 7 Issue III, Mar 2019- Available at www.ijraset.com

A Secure Access Policies Based Data Deduplication System

Dr. S. Hemalatha[1], Mrs. M. Vidhya[2], R. Ranjitha[3], T. Thendral[4], D. Vasumathi[5]
[1]*Professor,* [2]*Assistant Professor, Department of CSE, Panimalar Institute of Technology, Chennai, India*
[3, 4, 5]*Student, Department of CSE, Panimalar Institute of Technology, Chennai, India*

Abstract: ***Attribute-based encryption (ABE) has been widely used in cloud computing where a data provider outsources his/her encrypted data to a cloud service provider, and can share the data with users possessing specific credentials (or attributes). However, the standard ABE system does not support secure deduplication, which is crucial for eliminating duplicate copies of identical data in order to save storage space and network bandwidth. In this paper, we present an attribute-based storage system with secure deduplication in a hybrid cloud setting, where a private cloud is responsible for duplicate detection and a public cloud manages the storage. Compared with the prior data deduplication systems, our system has two advantages. Firstly, it can be used to confidentially share data with users by specifying access policies rather than sharing decryption keys. Secondly, it achieves the standard notion of semantic security for data confidentiality while existing systems only achieve it by defining a weaker security notion. In addition, we put forth a methodology to modify a ciphertext over one access policy into ciphertexts of the same plaintext but under other access policies without revealing the underlying plaintext. The main aim of this project is to achieve new distributed de-duplication systems. we present an attribute-based storage system with deduplication in a cloud setting with higher Trusty and security. In this Deduplication, techniques are most widely employed to backup data and minimize network and storage overhead by detecting and eliminating redundancy among data.***

I. INTRODUCTION

Hadoop is a framework that allows for the operating of large data set across a large number of computers with the use of programming models. It has been designed to scale up from single servers to thousands of computers, each has local storage. It makes Use of the artifact hardware. Hadoop is extremely Scalable and Fault Tolerant. Which provides resource management and programming for user applications and Hadoop Map cut back, which provides the programming model used to tackle largely distributed processing mapping knowledge and reducing it to a result. Big Data in most companies are processed by Hadoop by submitting the roles to Master. This type of usage is best-suited to highly scalable public cloud services; The Master distributes the job to its cluster and process map and reduces tasks sequentially. But nowadays the growing data need and the competition between Service Providers leads to the increased submission of jobs to the Master. This synchronal job submission on Hadoop forces the U.S. to try to programme on Hadoop Cluster so the time interval is going to be acceptable for every job.

II. LITERATURE SURVEY

Based on the previous search, Cloud Storage is recently as emerging topic in these eras. As the data are increasing, the storage become major issue for the people. There are different kind of Cloud Storage application such as One Drive, Sky Drive, Drop Box and Google Drive. Google Drive is gaining more popularity as it is user friendly than any other Cloud Storage Application. Google Drive is a Cloud Storage Application which allows user to store, share and edit the file in the cloud. In these paper, the authors will 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. From that the Authors will find, what type of data remnants can be found in user device. Motivated by the problem of avoiding duplication in storage systems, Bellare, Keelveedhi, and Ristenpart have recently put forward the notion of Message-Locked Encryption (MLE) schemes which sub- sumes convergent encryption and its variants. Such schemes do not rely on permanent secret keys, but rather encrypt messages using keys de- rived from the messages themselves. We strengthen the notions of security proposed by Bellare et al. by con- sidering plaintext distributions that may depend on the public parame-ters of the schemes. We refer to such inputs as lock-dependent messages. We construct two schemes that satisfy our new notions of security for message-locked encryption with lock-dependent messages. Our main construction deviates from the approach of Bell are et al. by avoiding the use of ciphertext components derived deterministically from the messages. We design a fully randomized scheme that supports an equality-testing algorithm dened on the ciphertexts. Our second construction has a deterministic ciphertext component that enables more .

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 7 Issue III, Mar 2019- Available at www.ijraset.com

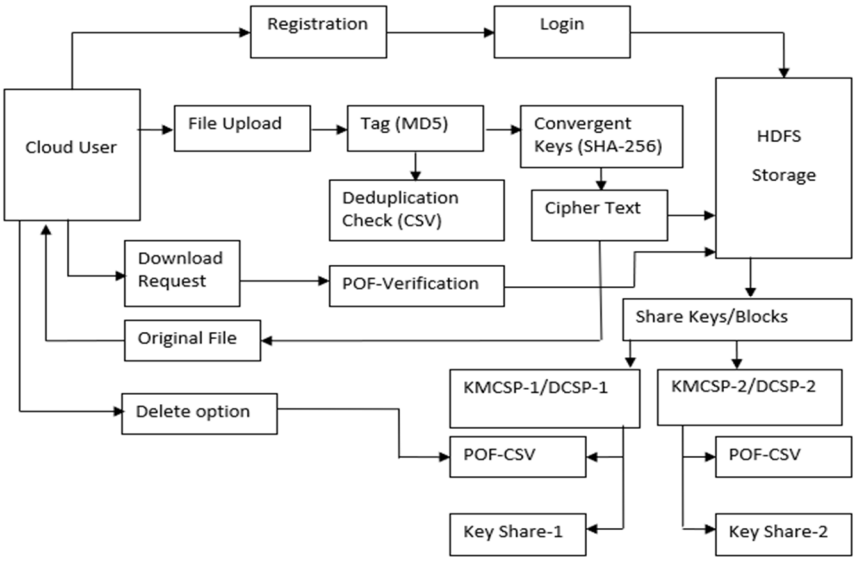
III. EXISTING SYSTEM

In the previous deduplication systems have only been maintain single-server setting. A data provider, Bob, intends to upload a file M to the cloud, and share M with users having certain credentials. In order to do so, Bob encrypts M under an access policy A over a set of attributes, and uploads the corresponding cipher text to the cloud, such that only users whose sets of attributes satisfying the access policy can decrypt the cipher text. Later, another data provider, Alice, uploads a cipher text for the same underlying file M but ascribed to a different access policy A0. Since the file is uploaded in an encrypted form, the cloud is not able to discern that the plaintext corresponding to Alice's cipher text is the same as that corresponding to Bob's, and will store M twice. Obviously, such duplicated storage wastes storage space and communication bandwidth. So they require that the deduplication storage systems provide reliability comparable to other high-available systems. However, as lots of deduplication systems and Storage systems are intended by users and applications for higher reliability, especially in archival storage systems where data are critical and should be preserved over long time periods. Cost increases to the storage of content as well as for the keys storage. Increase bandwidth with upload time.

IV. PROPOSED SYSTEM

In this paper, we have a tendency to gift the associate degree attribute-based storage system that employs ciphertext-policy attribute-based encryption (CP ABE) and supports secure deduplication. To enable the deduplication and distributed storage of the data across HDFS. And then using the two-way cloud in our storage system is built under a hybrid cloud architecture, where a private cloud calculate the computation and it manages the public cloud storage. The private cloud is provided with a trapdoor key associated with the corresponding ciphertext, with which it can transfer the ciphertext over one access policy into ciphertexts of the same plaintext under any other access policies while not being alert to the underlying plaintext. After receiving a storage request, the non-public cloud initial checks the validity of the uploaded item through the attached proof. If the proof is valid, the personal cloud runs a tag matching algorithmic rule to check whether identical knowledge underlying the ciphertext has been keep. If so, whenever it is necessary, it regenerates the ciphertext into a ciphertext of identical plaintext over associate access policy that is that the union set of each access policies. like the public cloud and private cloud. We have shown the idea of deduplication effectively and security is achieved by means that of Proof of Ownership of the file. That is attribute-based storage system ciphertext-policy attribute-based encryption (CP-ABE) and supports secure deduplication

V. ARCHITEHTURE DIAGRAM



In the architecture diagram cloud user has been registered and login to the cloud storage. The file is uploaded and tagged by using MD5, keys are generated using SHA-256 and stored in HDFS storage. The file undergoes duplication check(CSV) with the original file. If the file is matched with original file the download request has been given followed by POF- verification then the file is accessed from the HDFS storage. During the duplication check existing data can be eliminated using Delete option.

©IJRASET: All Rights are Reserved

VI. MODULES

A. Cloud user Registration

In this module cloud user first register the user details (Name, password, email, access policy, mobile no, dob). And then log in the user credential details like username, password[5]. Once username and password are valid open the user profile screen will be displayed.

B. File Upload With Access Policies

In this module, the User will choose the file and uploads to Storage where the HDFS storage system. In the system will generate a signature in a particular file and then split into multiple blocks. Each block will be generating a signature with the key[6]. In the signature by using the MD5 message-digest algorithm is cryptographic hash function producing a 128-bit hash value typically expressed in text format as a 32 digit hex value so that files of the same are deduplicated. After that generate convergent keys for each block splitting to store CSV file like filename, file path, blocks, username, password and block keys. Encrypt the blocks by the RSA algorithm is asymmetric cryptography algorithm. Asymmetric really means it works on 2 totally different keys i.e. Public Key and Private Key. As the name describes that the general public secret is given to everybody non-public|and the personal} secret is unbroken private. Here the plain text is encryption to ciphertext and stored in a slave system. Blocks are stored in Distributed HDFS Storage Providers. After uploading the file to set the access policies with a set security question

.

C. Detection Deduplication method

File-level knowledge deduplication compares a file to be secured or archived with copies that area unit already holds on. This is done by checking its attributes against the Associate in Nursing index. If the file is exclusive, it is stored and the index is updated; if not, 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 signature match checking looks within a file and saves unique iterations of each block. All the blocks area unit was broken into chunks with identical mounted length. Each chunk of information is processed employing a hash algorithmic rule like MD5 or SHA-1.

D. Download User File

The final model user requests for downloading their own document which they have uploaded in HDFS storage. In this download request will analysis the user attribute once it will match then ask the security questions for a particular file. After complete, the process needs proper ownership verification. After 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

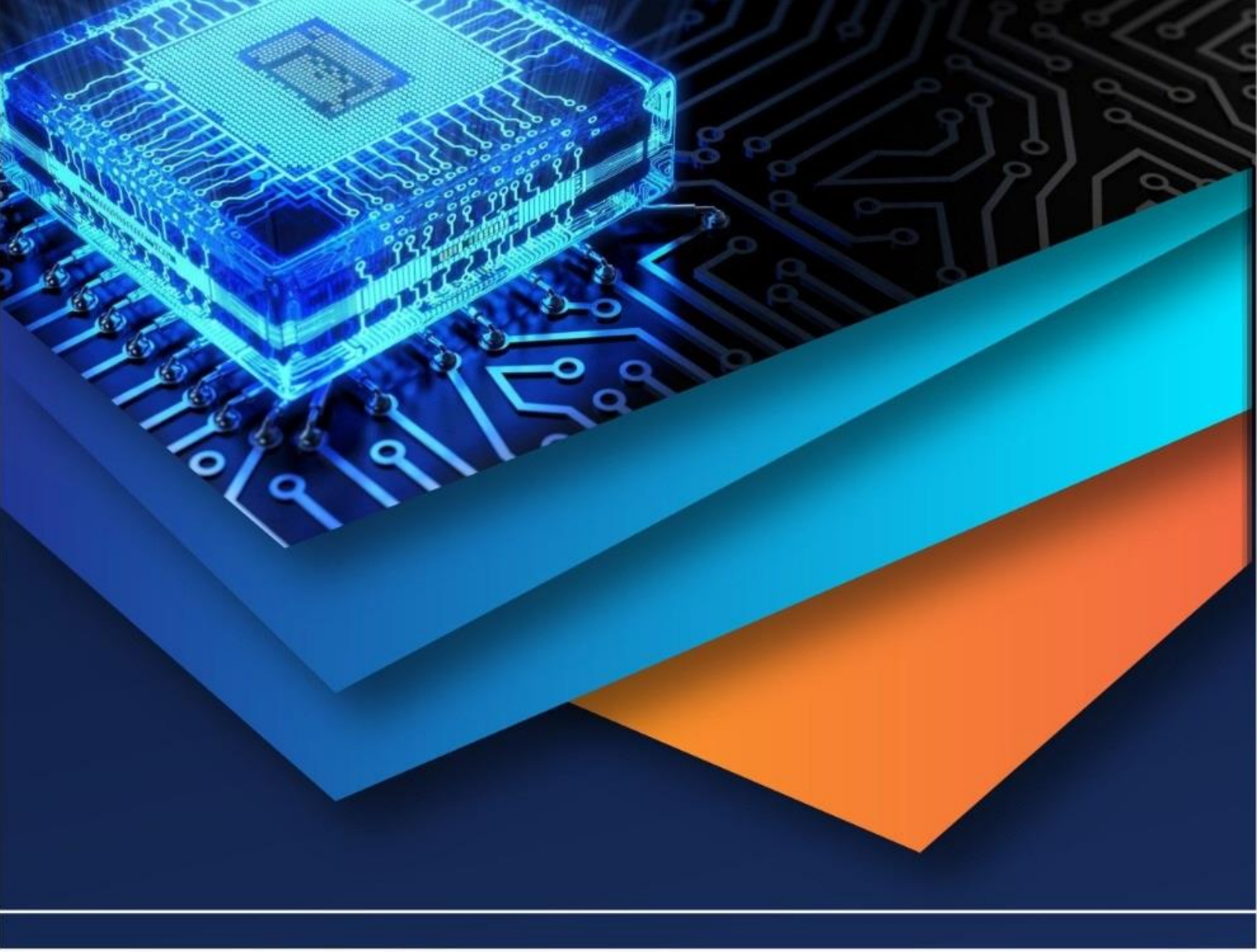
VII. CONCLUSION

In this project, the new distributed de-duplication systems with file-level and fine-grained block-level data deduplication, higher reliability in which the data chunks are distributed across HDFS storage, reliable key management in secure de-duplication and the security of tag consistency and integrity were achieved.

REFERENCE

[1] S. Thavalengal and P. Corcoran, "User authentication on smartphones: that specialize in iris biometry," IEEE client physics Magazine, vol. 5, no. 2, pp. 87–93, 2016
[2] M. De Marsico, M. Nappi, and D. Riccio, "Noisy iris recognition integrated theme, "PatternRecognitionLetters,vol.33,no.8,pp.1006– 1011, 2012.
[3] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in Advances in Neural scientific discipline Systems, 2012, pp. 1097–1105.
[4] A. Gangwar and A. Joshi, "DeepIrisNet: Deep iris representation with applications in iris recognition and cross-sensor iris recognition," in International Conference on Image Processing, 2016, pp. 2301–2305.
[5] H. Proenc¸a and J. C. Neves, "IRINA: Iris recognition (Even) in inaccurately segmented data," in IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 538–547.
[6] Z. Zhao and A. Kumar, "Towards more accurate iris recognition using deeply learned spatially corresponding features," in International Conference on Computer Vision, 2017, pp. 22–29.

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)