IARJSET

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology
Vol. 8, Issue 4, April 2021

DOI: 10.17148/IARJSET.2021.8410

Secure Encrypted Data with Authorized Deduplication in Cloud

Mr.G.Deeban Chakkarawarthi[1], Abinaya B[2], Gayathiri K[3], Kencey K[4], Vanitha M[5]

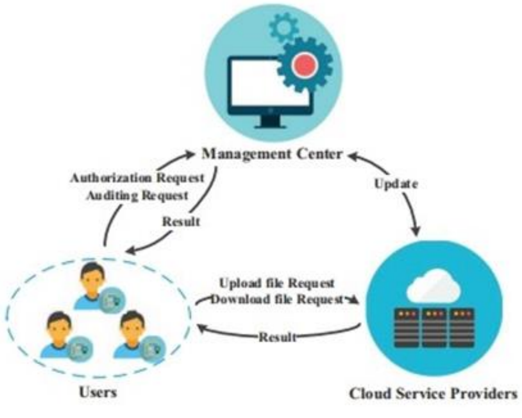
Department of Computer Science and Engineering,
JCT College of Engineering and Technology Coimbatore, TamilNadu, India[1,2,3,4,5]

Abstract— The basic notion of deduplication is storage of duplicate data only for a single time. Thus, a user willing to upload a stored file will have to be first added by the cloud provider in the owner list for that particular file. This is the reason why deduplication has been rapidly adopted by various providers of cloud storage. Today, it has become a popular approach for minimizing storage space and for uploading bandwidth and assists largely in increasing the scalability of data. Deduplication also eliminates the fear of surplus data by maintaining a single physical copy and refers any surplus data to this copy and is the best alternative of multiple data copies having the same data.

Keywords— Encryption, De-Duplication, Convergent Encryption.

I. INTRODUCTION

Big data is a collection of massive and complex data sets and data volume that include the huge quantities of data, data management capabilities, social media analytics and real-time data. Big data analytics is the process of examining large amounts of data. There exist large amounts of heterogeneous digital data. Big data is about data volume and large data set's measured in terms of terabytes or petabytes. This phenomenon is called Bigdata. After examining of Bigdata, the data has been launched as Big Data analytics. In this paper, presenting the 5Vs characteristics of big data and the technique and technology used to handle big data. The rapid development of cloud computing and big data technology changes user's method and efficiency in processing information, the cloud servers provide the scalable computing and efficient storage to users in anytime and anywhere. The rising popularity of cloud computing can be attributed to lower costs, easily usable processing resources and increased storage. There has also been an unexpected rise in the use of online digital data which multiplies the significance of cloud storage for effective costing and better utilization of power. With increasing volume in data, the Total Cost of Ownership (TCO) is also increasing including human organization, management and storage setup cost.



II. PROBLEM ANALYSIS

Protect the confidentiality of sensitive data while supporting deduplication, Cloud computing provides seemingly unlimited "virtualized" resources to users as services across the whole Internet, while hiding platform and implementation details. In il the hash-based deduplication schemes, cryptographic
hash functions, such as the MD and SHA family, are used for calculating chunk fingerprints due to their very low probability of hash collisions that renders data loss extremely unlikely.

IARJSET

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Vol. 8, Issue 4, April 2021

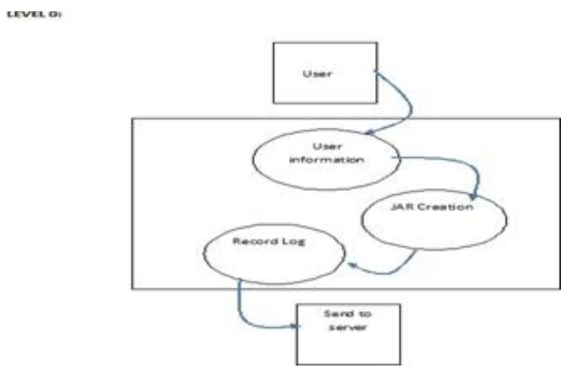
DOI: 10.17148/IARJSET.2021.8410

III. RELATED WORK

In order to tackle the problem that the unauthorized users can access the user information only by supplying the hash value of the file. Halevi et al. [1] proposed the proof of ownership (PoW), which is an interaction protocol between client side and server side to verify the ownership of that client.[2] proposed a private data deduplication in data storage, where a client held a private data proves to a server stored a summary string of the data that he/she is the owner of that data without revealing further information to the server.[3] proposed a cryptographic primitive to enhance the security of client-side deduplication in the bounded leakage setting where a certain amount of efficiently-extractable information about file F is leaked.[4] proposed a s-PoW scheme, which requests some particular random bits of the file from the verified client. [5] proposed a new deduplication scheme with multimedia data, which is based on randomized convergent encryption and privilege-based encryption to achieve authorized deduplication and user revocation. [6] proposed a secure encrypted data deduplication scheme, which exploits the homorphic encryption algorithm to achieve security data deduplication [7], and supports ownership check and user revocation.[8] proposed a secure data deduplication scheme based on CP AES algorithm [9] in cloud. [10] proposed a novel data deduplication scheme and gave a concept of hybrid cloud environment, where the generation of encryption keys is related to the corresponding privilege, the private cloud server is responsible for management and storage of the user's keys, and the public cloud server stores the ciphertext and performs the data deduplication.

IV. PROPOSED SYSTEM

The convergent encryption technique has been proposed to encrypt the data before outsourcing. To better protect data security, this paper makes the first attempt to formally address the problem of authorized data deduplication. Different from traditional deduplication systems, the differential privileges of users are further considered in duplicate check besides the data itself. We also present several new deduplication constructions supporting authorized duplicate check in hybrid cloud architecture.



Security analysis demonstrates that our scheme is secure in the definitions specified in the proposed security model. As a proof of concept, we implement a prototype of our proposed authorized duplicate check scheme and conduct test bed experiments using our prototype. We show that our proposed authorized duplicate check scheme incurs minimal overhead compared to normal operations. The 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 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.

AES Algorithm

The Advanced Encryption Standard (AES) is an encryption algorithm for securing sensitive but unclassified material. Implementations of all of the above were tested extensively in ANSI C and Java languages for speed and reliability in such measures as encryption and decryption speeds, key and algorithm set-up time and resistance to various attacks, both in hardware- and software-centric systems. Once again, detailed analysis was provided by the global cryptographic community.

AES is based on a design principle known as a Substitution permutation network. It is fast in both software and hardware. Unlike its predecessor, DES, AES does not use a Feistel network.AES has a fixed block size of 128 bits and a key size of 128, 192, or 256bits, whereas Rijndael can be specified with block and key sizes in any multiple of 32bits, with a minimum of 128bits. The block size has a maximum of 256bits, but the key size has no theoretical maximums operates on a 4×4 column-major order matrix of bytes, termed the *state* (versions of Rijndael with a larger block size have additional columns in the state). Most AES calculations are done in a special finite field. The AES cipher is specified as a number of repetitions of transformation rounds that convert the input plaintext into the final output of cipher text. Each round consists of several processing steps, including one that depends on the encryption key. A set of reverse rounds are applied to transform cipher text back into the original plaintext using the same encryption key.

Key Expansion—round keys are derived from the cipher key using Rijndael's key schedule
Initial Round
Add Round Key—each byte of the state is combined with the round key using bitwise xor
Rounds
SubBytes—a non-linear substitution step where each byte is replaced with another according to a lookup table.
ShiftRows—a transposition step where each row of the state is shifted cyclically a certain number of steps.
MixColumns—a mixing operation which operates on the columns of the state, combining the four bytes in each column.
AddRoundKey
Final Round (no MixColumns) SubBytes
ShiftRows AddRoundKey
A block cipher key is fixed for each of the currently allowed key sizes, i.e., AES-128, AES-192, AES-256, two key TDEA and three key TDEA. For each key, the generation of the associated sub keys is given, followed by four examples of MAC generation with the key. The messages in each set of examples are derived by truncating a common fixed string of 64 bytes. All strings are represented in hexadecimal notation, with a space (or a new line) inserted every 8 symbols, for readability. As in the body of the Recommendation, K1 and K2denote the sub keys, M denotes the message, and T denotes the MAC. For the AES algorithm examples, ten is 128, i.e., 32 hexadecimal symbols, and K denotes the key.
For the TDEA examples, ten is 64, i.e.,16 hexadecimal symbols, and the key, K, is the ordered triple of strings, (Key1, Key2, Key3). For two key TDEA, Key1 = Key3. D.1 AES-128
For Examples 1–4 below, the block cipher is the AES algorithm with the following 128 bits key: K 2b7e1516 28aed2a6 abf71588 09cf4f3c. Subkey Generation
CIPHK (0128)7df76b0c 1ab899b3 3e42f047 b91b546f K1 fbeed618 35713366 7c85e08f 7236a8de K2 f7ddac30 6ae266cc f90bc11e e46d513b

V. SYSTEM IMPLEMENTATION

The proposed system implementation consists of four different modules as the following, cloud user registration, File upload and access policies, Deduplication method & Download user files.

A. CLOUD USER REGISTRATION

The cloud user will first register the user details such as Name, password, email, mobile number. we can then login with credential details like username, password. Once user name and password is valid open the user profile screen will be displayed.

B. FILE UPLOAD WITH ACCESS POLICIES

The registered user can then have the privilege to access and upload any file or document to the cloud. The upload file is splitted, each splitted file is tagged and a spontaneous product key is generated and is stored in the cloud with reference to the file that has been uploaded. Encrypt the blocks by AES algorithm is asymmetric cryptography algorithm. Asymmetric actually means that it works on two different keys i.e., Public Key and Private Key. 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.

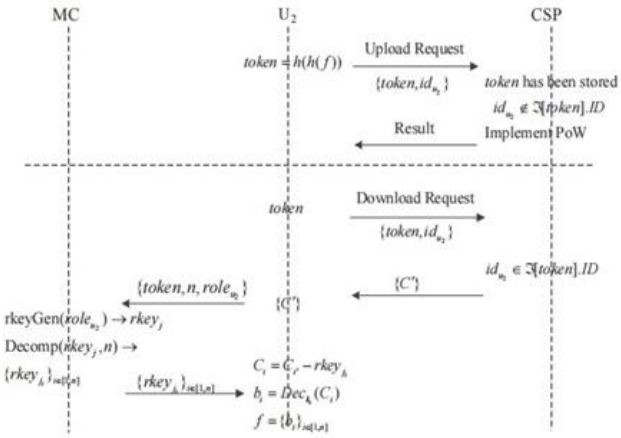
C. FILE LEVEL DEDUPLICATION

File-level data deduplication compares a file to be backed up or archived with copies that are already stored. This is done by checking its attributes against an index. If the file is unique, it is stored and the index is updated 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.

D. USER FILE DOWNLOAD MODULE

The final model user request for downloading their own document which they have uploaded in 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 registered user can upload, Download and Delete the files that he has contributed. This allows the user to have access to the cloud environment. The uploaded file or document is stored in the cloud using the AES Algorithm, so that the file has been encrypted and has been assigned a new or unique checksum and secret key at the time of storage in the cloud.

IARJSET
ISSN (Online) 2393-8021
ISSN (Print) 2394-1588
International Advanced Research Journal in Science, Engineering and Technology
Vol. 8, Issue 4, April 2021
DOI: 10.17148/IARJSET.2021.8410



VI. CONCLUSION

we proposed a complete system to securely outsource log records to a cloud provider. We reviewed existing solutions and identified problems in the current operating system-based logging services such as syslog and practical difficulties in some of the existing secure logging techniques. In this work, find out the challenges for a secure cloud-based log management service

Client-Side Deduplication eliminate duplicate which results in effective utilization of the resources such as storage space and bandwidth consumption instead of transmitting same data repeatedly. Deduplication has benefits like reduced infrastructure costs, management costs and reduced downtime. By using convergent encryption and Merkle based deduplication can be done in a secured and efficient way.

VII. FUTURE ENHANCEMENT

We propose SecCloud and SecCloud+ in Aiming at achieving both data integrity and deduplication in cloud. SecCloud introduces an auditing entity with maintenance of a Map Reduce cloud, which helps clients generate data tags before uploading aSecurity analysis demonstrates that our scheme is secure in terms of the definitions specified in the proposed security model. As a proof of concept, we implement a prototype of our proposed authorized duplicate check scheme and conduct test bed experiments using our prototype. We show that our proposed authorized duplicate check scheme incurs minimal overhead compared to normal operations. Security analysis demonstrates that our scheme is secure in terms of the definitions specified in the proposed security model. As a proof of concept, we implement a prototype of our proposed authorized duplicate check scheme and conduct test bed experiments using our prototype. We show that our proposed authorized duplicate check scheme incurs minimal overhead compared to normal operations well as audit the integrity of data having been stored in cloud. In addition, SecCloud enables secure deduplication through introducing a Proof of ownership protocol and preventing the leakage of side channel information in data deduplication. Compared with previous work, the computation by user in SecCloud is greatly reduced during the file uploading and auditing phases. SecCloud+ is an advanced construction motivated by the fact that customers always want to encrypt their data before uploading and allows for integrity auditing and secure deduplication directly on encrypt data.

VIII. REFERENCES

[1] S. Halevi, D. Harnik, B. Pinkas, and A. Shulman-Peleg, “Proofs of ownership in remote storage systems,” in Proceedings of the 18th ACM SIGSAC Conference on Computer and Communications Security. ACM, 2011, pp. 491–500.

[2] W. K. Ng, Y. Wen, and H. Zhu, “Private data deduplication protocols in cloud storage,” in Proceedings of the 27th Annual ACM Symposium on Applied Computing. ACM, 2012, pp. 441–446.

[3] J. Xu, E.-C. Chang, and J. Zhou, “Weak leakage-resilient client-side deduplication of encrypted data in cloud storage,” in Proceedings of the 8th ACM SIGSAC symposium on Information, computer and communications security. ACM, 2013, pp. 195–2016.

[4] R. Di Pietro and A. Sorniotti, “Boosting effificiency and security in proof of ownership for deduplication,” in Proceedings of the 7th ACM Symposium on Information, Computer and Communications Security. ACM, 2012, pp. 81–82.

[5] H. Kwon, C. Hahn, D. Kim, and J. Hur, “Secure deduplication for multimedia data with user revocation in cloud storage,” Multimedia Tools and Applications, vol. 76, no. 4, pp. 5889–5903.

[6] W. Ding, Z. Yan, and R. H. Deng, “Secure encrypted data deduplication with ownership proof and user revocation,” in International Conference on Algorithms and Architectures for Parallel Processing. Springer, 2017, pp. 297–312.

DOI: 10.17148/IARJSET.2021.8410

[7] X. Liu, R. H. Deng, W. Ding, R. Lu, and B. Qin, "Privacy-preserving outsourced calculation of flfloating point numbers," in IEEE Transactions on Information Forensics and Security.

[8] H. Tang, Y. Cui, C. Guan, J. Wu, J. Weng, and K. Ren, "Enabling ciphertext deduplication for secure cloud storage and access control," in Proceedings of the 11th ACM on Asia Conference on Computer and Communications Security. ACM, 2016, pp. 59–70.

[9] Q. Li, J. Ma, R. Li et al., "Secure, effificient and revocable multi-authority access control system in cloud storage," Computers & Security, vol. 59, no. 6, pp. 45–59, 2016.

[10] J. Li, Y. K. Li, X. Chen, P. Lee, and W. Lou, "A hybrid cloud approach for secure authorized deduplication," IEEE Transactions on Parallel and Distributed Systems, vol. 26, no. 5, pp. 1206–1216.