ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

A Cloud Approach for Secure Data Deduplication

M. Palani [1], R. Sankar Ramkumar [2], P. Velpandi [3] and R. Kanagaselvi [4]

U.G Students, Dept. of CSE, Sree Sowdambika College of Engineering, Chettikurichi, Aruppukottai, Virudhunagar, Tamil Nadu ,India [1,2,3]

Assistant Professor, Dept. of CSE, Sree Sowdambika College of Engineering, Chettikurichi, Aruppukottai, Virudhunagar, Tamil Nadu, India [4]

ABSTRACT: Data management adaptable in distributed computing, duplication has been a surely understood strategy and has pulled in more consideration as of late. Rather than keeping different data duplicates with the same substance, duplication takes out repetitive data by keeping one and only physical duplicate and pertaining other excess data to that duplicate. Convergent encryption has been proposed to implement data privacy while making duplication possible. It scrambles/decodes a data duplicate with a convergent key, which is gotten by processing the cryptographic hash estimation of the substance of the data. To permit the cloud to perform duplication on the figure writings and the evidence of proprietorship keeps the unapproved client to get to the document.

KEYWORDS: Data management, duplicate data, distributed computing, Key generation algorithm and AES.

I. INTRODUCTION

Distributed computing gives apparently boundless "virtualized" assets to clients as administrations over the entire Internet, while hiding platform and implementation elements. Today's cloud administration suppliers offer both exceptionally accessible capacity and hugely parallel computing assets at moderately low expenses. As distributed computing gets to be common, an expanding measure of data is being storing in the cloud and imparted by clients to indicated benefits, which characterize the privileges of the stored data. One basic test of distributed storage administrations is the administration of the constantly expanding volume of data. To make data administration adaptable in distributed computing, duplication [1,2] has been a surely understood procedure and has pulled in more consideration as of late. Data duplication is a specific data compression technique for dispensing with copy duplicates of rehashing data away. The method is utilized to enhance storage usage and can likewise be connected to network data exchanges to diminish the quantity of bytes that should be sent. [3-5]. Rather than keeping different data duplicates with the same substance, duplication disposes of repetitive data by keeping one and only physical duplicate and alluding other excess data to that duplicate. Reduplication can occur at either the file level or the block level. For File-level duplication, it dispenses with copy duplicates of the same document. Reduplication can likewise happen at the block level, which dispenses with duplicate blocks of data that happen in non-indistinguishable files.

Despite the fact that data duplication brings a great deal of advantages, security and protection concerns emerge as clients' sensitive data are helpless to both insider and outsider attacks. Conventional encryption, while giving data privacy, is inconsistent with data duplication. In particular, customary encryption requires distinctive clients to encode their data with their own keys. Accordingly, indistinguishable data duplicates of various clients will prompt different cipher texts, making duplication impossible. Convergent encryption [6,9] has been proposed to authorize data privacy while making duplication feasible. It encrypts/decrypts a data duplicate with a convergent key, which is acquired by processing the cryptographic hash estimation of the content of the data duplicate. After key generation and data encryption, clients hold the keys and send the cipher text to the cloud. Since the encryption operation is deterministic and is derived from the data content, indistinguishable data duplicates will create the same convergent key and thus the same cipher text. To avoid unapproved access, a secure proof of ownership (POW) protocol [7,8] is likewise expected to give the verification that the client to be sure possesses the same file when a copy is found. After the confirmation,

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

ensuing clients with the same document will be given a pointer from the server without expecting to transfer the same file. A client can download the encrypted file with the pointer from the server, which must be unscrambled by the comparing data proprietors with their convergent keys. In this manner, convergent encryption permits the cloud to perform duplication on the cipher texts and the verification of ownership keeps the unapproved client to access the file. [10].

II. PROBLEM STATEMENT

1. Data duplication frameworks, the private cloud is included as an intermediary to permit data owner/clients to safely perform copy check with differential benefits.
2. This architecture is real time and has pulled in much consideration from researchers.
3. The data owners just outsource their data utilizing so as to stockpile open cloud while the data operation is maintained in private cloud.
4. Conventional encryption, while giving data confidentiality, is incongruent with data duplication.
5. Identical data duplicates of various clients will prompt different cipher texts, making duplication unimaginable.

III. PROPOSED SYSTEM

We improve our framework in security. In particular, we introduce an advanced scheme to bolster stronger security by encoding the file with differential privilege keys. Along these lines, the clients without comparing benefits can't perform the copy check. Besides, such unauthorized clients can't decrypt the cipher text even connive with the S-CSP. Security analysis exhibits that our framework is secure as far as the definitions determined in the proposed security model.

3.1. Key Generation Algorithm

Key generation is the procedure of creating keys in cryptography. A key is utilized to encode and decode whatever data is being encoded/decoded. Present day cryptographic frameworks incorporate symmetric-key algorithms, (for example, DES and AES) and public key algorithms, (for example, RSA). Symmetric-key algorithms utilize a solitary shared key; keeping data secret requires maintaining this key secret. Public key algorithms utilize a public key and a private key. The public key is made accessible to anybody (frequently by method for a computerized declaration). A sender encodes data with public key; just the holder of the private key can decode this data. Since public key algorithms have a tendency to be much slower than symmetric-key algorithms, advanced frameworks, for example, TLS and SSH utilize a blend of the two: one gathering gets the other's public key, and encodes a little bit of data (either a symmetric key or some data used to produce it). The rest of the discussion utilizes a (regularly quicker) symmetric-key algorithm for encryption.

3.2. Symmetric-key algorithms

Symmetric-key algorithms [1] are algorithms for cryptography that utilization the same cryptographic keys for both encryption of plaintext and decryption of cipher text. The keys might be indistinguishable or there might be a basic change to go between the two keys. The keys, by and by, constitute to a mutual secret between two or more gatherings that can be utilized to keep up a private data link.[2] This prerequisite that both sides have admittance to the secret key is one of the fundamental drawbacks of symmetric key encryption, in contrast with public key encryption (otherwise called asymmetric key encryption).[3].

3.3. AES algorithm in cipher block chaining

In cryptography, a method of operation is a algorithm that uses a block cipher to give a data administration, for example, confidentiality or authenticity.[11] A block cipher by itself is desirable for the protected cryptographic change (encryption or decryption) of one fixed length group of bits called a block.[12] A method of operation depicts how to over and over apply a cipher's single-block operation to safely change measures of data bigger than a block.[13-15] Most modes require a novel binary sequence, regularly called an initialization vector (IV), for every encryption operation. The IV must be non-rehashing and, for a few modes, irregular too. The instatement vector is utilized to guarantee unmistakable cipher texts are created notwithstanding when the same plaintext is encoded various times freely with the same key.[6] Block cipher have one or more block size(s), yet during change the block size is constantly

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

settled. Block cipher modes work on entire blocks and require that the last part of the data be padded to a full block in the event that it is smaller than the present block size.[2] There are, be that as it may, modes that don't require padding since they successfully utilize a block cipher as a stream cipher.

3.4. Advantages

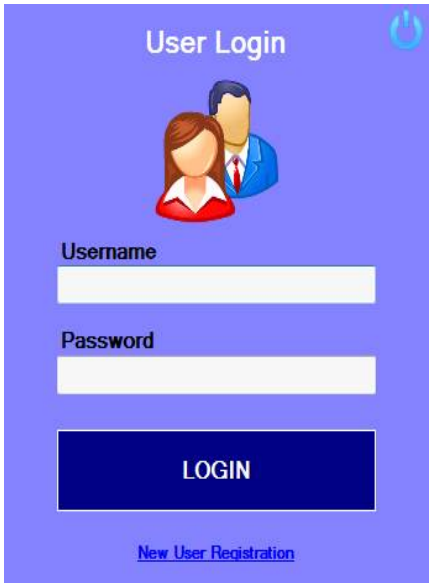
- The client is just permitted to perform the copy check for files set apart with the comparing benefits.
- We present an advanced scheme to bolster stronger security by encoding the file with differential benefit keys.
- Reduce the capacity size of the labels for integrity check. To improve the security of reduplications and ensure the data confidentiality.

IV. IMPLEMENTATION

- ✓ User Management
- ✓ File Validation and Schema Key
- ✓ Private Key Generation and Implementing Cipher Text
- ✓ File Encryption and Decryption
- ✓ Performance Analysis

4.1. User Management

Client can register in the administration supplier and stored the legitimate file in private cloud server. Private Server project is utilized to show the private cloud which deals with the private keys and handles the file token algorithm. A Storage Server program is utilized to demonstrate the S-CSP which stores and copies files.



4.2. File Validation and Schema Key

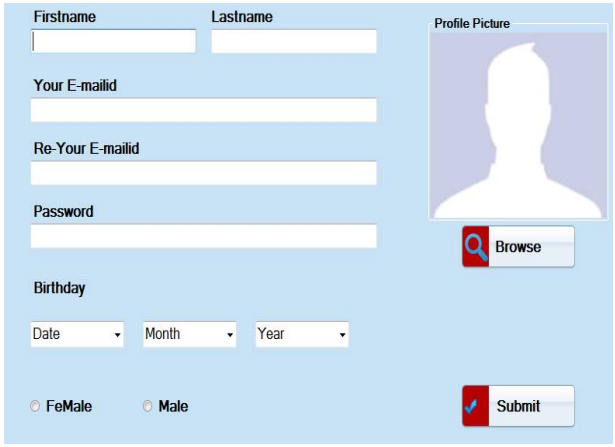
A file copy is found, the client needs to run the POW protocol POW with the S-CSP to demonstrate the file ownership. In the event that the verification is passed, the client will be given a pointer to the file. A proof from the S-CSP will be returned, which could be a mark. The tag of a file F will be controlled by the file F and the benefit. To demonstrate the distinction with conventional notation of label, we call it file token.

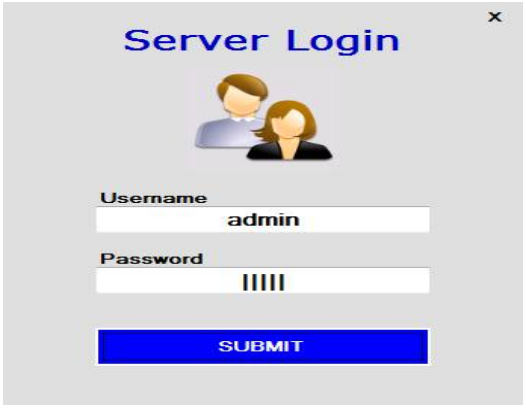
ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016





4.3. Private Key Generation and Implementing Cipher Text

We will utilize the hash functions to characterize the label generation functions and merged keys in this segment. Encryption, to bolster copy check, the key is gotten from the document F by utilizing some cryptographic hash functions. To stay away from the deterministic key generation, the encryption key for file F in our framework will be created with the guide of the private key cloud server with benefit key KP. The Cipher text C and the merged key K as inputs and outputs yields the master data duplicate.

4.4. File Encryption and Decryption

Every file is ensured with the convergent encryption key and privilege keys to understand the approved reduplication with differential benefits. Taking into account this suspicion, we demonstrate that frameworks are secure as for the accompanying security investigation. Encryption permits the cloud to perform reduplications on the cipher texts and the ownership prevents keeps the unapproved client to get to the file. The security of data in our first development could be ensured under this security thought. Decrypted by comparing data owners with their convergent keys are analyzed.

V. PERFORMANCE ANALYSIS

Random Key Generation is utilized in between the Symmetric encryption algorithm to confirm the client, in which the client needs to enter an arbitrary key number. Cryptography algorithm Symmetric encryption strategy is utilized to enhance the security and key level. The data is finally stored in the cloud server and measure the storage. All the binary codes are decoded to text file and the text file is unfastened to uncover the data file.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

VI. CONCLUSION

In this paper, the idea of approved data duplication was proposed to ensure the data security by including differential benefits of clients in the copy check. We likewise displayed a few new duplication developments supporting approved copy check in hybrid cloud architecture, in which the copy check tokens of documents are produced by the private cloud server with private keys. Security analysis shows that our approaches are secure in terms of insider and outsider attacks determined in the proposed security model. As a proof of idea, we actualized a model of our proposed approved copy check plan and direct test bed investigates our model. We demonstrated that our approved copy check plan brings about insignificant overhead contrasted with united encryption and network transfer.

REFERENCES

[1] OpenSSL Project, (1998). [Online]. Available: http://www. openssl.org/
[2] P. Anderson and L. Zhang, "Fast and secure laptop backups with encrypted de-duplication," in Proc. 24th Int. Conf. Large Installation Syst. Admin., 2010, pp. 29–40.
[3] M. Bellare, S. Keelveedhi, and T. Ristenpart, "Dupless: Serveraided encryption for deduplicated storage," in Proc. 22nd USENIX Conf. Sec. Symp., 2013, pp. 179–194.
[4] M. Bellare, S. Keelveedhi, and T. Ristenpart, "Message-locked encryption and secure deduplication," in Proc. 32nd Annu. Int. Conf. Theory Appl. Cryptographic Techn., 2013, pp. 296–312.
[5] M. Bellare, C. Namprempre, and G. Neven, "Security proofs for identity-based identification and signature schemes," J. Cryptol., vol. 22, no. 1, pp. 1–61, 2009.
[6] M. Bellare and A. Palacio, "Gq and schnorr identification schemes: Proofs of security against impersonation under active and concurrent attacks," in Proc. 22nd Annu. Int. Cryptol. Conf. Adv. Cryptol., 2002, pp. 162–177.
[7] S. Bugiel, S. Nurnberger, A. Sadeghi, and T. Schneider, "Twin clouds: An architecture for secure cloud computing," in Proc. Workshop Cryptography Security Clouds, 2011, pp. 32–44.
[8] J. R. Douceur, A. Adya, W. J. Bolosky, D. Simon, and M. Theimer, "Reclaiming space from duplicate files in a serverless distributed file system," in Proc. Int. Conf. Distrib. Comput. Syst., 2002, pp. 617–624
. [9] D. Ferraiolo and R. Kuhn, "Role-based access controls, " in Proc. 15th NIST-NCSC Nat. Comput. Security Conf., 1992, pp. 554–563.
[10] GNU Libmicrohttpd, (2012). [Online]. Available: http://www. gnu.org/software/libmicrohttpd/
[11] S. Halevi, D. Harnik, B. Pinkas, and A. Shulman-Peleg, "Proofs of ownership in remote storage systems," in Proc. ACM Conf. Comput. Commun. Security, 2011, pp. 491–500.
[12] J. Li, X. Chen, M. Li, J. Li, P. Lee, and W. Lou, "Secure deduplication with efficient and reliable convergent key management," in Proc. IEEE Trans. Parallel Distrib. Syst., http:// doi.ieeecomputersociety.org/10.1109/TPDS.2013.284, 2013.
[13] libcurl, (1997). [Online]. Available: http://curl.haxx.se/libcurl/
[14] C. Ng and P. Lee, "Revdedup: A reverse deduplication storage system optimized for reads to latest backups," in Proc. 4th AsiaPacific Workshop Syst., http://doi.acm.org/10.1145/2500727. 2500731, Apr. 2013.
[15] W. K. Ng, Y. Wen, and H. Zhu, "Private data deduplication protocols in cloud storage," in Proc. 27th Annu. ACM Symp. Appl. Comput., 2012, pp. 441–446