International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2020 IJSRCSEIT | Volume 6 | Issue 2 | ISSN : 2456-3307
DOI : https://doi.org/10.32628/CSEIT2062159

Authorized Deduplication of Encrypted data in Cloud

Prof. Milind B. Waghmare, Suhasini V. Padwekar
Computer Science and Engineering, Government College of Engineering, Amravati , Maharashtra, India

ABSTRACT

Cloud computing technology is rapidly developing nowadays. The number of files stored and processed is increasing per day. This increase brings severe challenge in requirement of space, processing power and bandwidth. More than half of the data generated in the cloud is duplicate data. To handle this data, deduplication technique is used which eliminates duplicate copies of data. This removal of duplicate data increases storage efficiency and reduce cost. In this paper, we propose secure role re-encryption system which allows authorized deduplication of data and also maintains privacy of data. This system is based on convergent algorithm and re-encryption algorithm that encrypts the user data and assign role keys to each user. This system grants privileges to users in order to maintain ownership of each user so that authorized users can access the data efficiently. In this system management center is introduced where the file is being encrypted and role keys are generated to handle authorized requests. Role keys are stored in Merkle hash tree which maps relationship between roles and keys. Authorized user who has particular role-encryption key can access the file. Convergent algorithm and role re-encryption algorithm allows access of specific file without leakage of private data. Dynamic updating of user privileges is achieved.

Keywords : Authorized user, Cloud computing, Data Deduplication, Data privacy, Confidentiality

I. INTRODUCTION

Cloud computing technology has emerged rapidly in various areas. This technology provides huge storage space, servers, applications and processing power capacity. So, the users and enterprises choose to store their sensitive data on the cloud. The cloud computing technology has evolved with many rapid developments. The data stored on the cloud is increasing daya by day. According to report thousands of users sign in to Facebook accounts per minute. Hundreds of video are uploaded on YouTube per minute. On daily basis millions of search request of users are forwarded to Google and Instagram users put up millions of photos in one minute [1], [12], and [13]. The data uploaded on cloud regularly is increasing with high speed. The data has been constantly increasing, 1.8 ZB in 2011, 4.4 ZB in 2013, 8.61 in 2015 and expected to get 44 ZB by the end of 2020. Here it is observed that degree of expansion is almost half more than the previous year [1], [14]. It is being observed that half of the data stored on the cloud is duplicate data. The cost of maintaining such data is very high as it requires huge processing power, space and expenditure. The cost of maintaining duplicate data is 8 times more than that of processing and storing original data [1], [15]. The increase in upload of data on the cloud results in huge maintenance cost which is problem to be solved. Reduction in management expenditure and improving storage efficiency is a critical problem.

To tackle above situation data deduplication technique can be implemented. Data deduplication technique is one of the compression techniques which eliminate redundant data on the cloud. This technique is widely used by cloud service providers. Duplicate copies of data are removed and only single instance of data is maintained on the cloud. This helps in improving storage utilization and reduces expenditure cost. The data uploaded on the cloud by the user have private or sensitive information. As user uploads data on the cloud it loses its control over the data. Cloud service provider is considered curious which risks leakage of private data of the user. CSP performs deduplication to reduce duplicate data, the foe may try to steal users sensitive data.

To deal with the above issues secure role re-encryption algorithm can be used. This algorithm is based on CE and role re-encryption algorithm to accomplish authorized data deduplication [1]. The main objectives of this scheme are as follows:

- This system prevents emitting of privacy data by performing authorized data deduplication.
- The system uses convergent encryption algorithm to protect data confidentiality, utilizes the role re-encryption algorithm for authorized access only.
- Management center is introduced to deal with user request and generate the role re-encryption key which allows particular user with that key to access particular file.

The rest of this paper is organized as follows. In Section II, the data classification based on processing unit and data execution object is presented. In Section III we have given Literature survey. In Section IV, system model for data deduplication is discussed. Consequently, we compare techniques for data deduplication in tabular form in Section V. Finally, the conclusion is given in Section VI.

II. LITERATURE SURVEY

In [1] the authors have introduced SRRS system which comprise of convergent algorithm to maintain data confidentiality and used role re-encryption algorithm to accomplish authorized data deduplication effectively. Management center is introduced to manage keys and user's roles. With the introduction of management center in the system, computational cost and overhead gets reduced on the client side. The SRRS system performs data deduplication and reduces storage space requirement and bandwidth consumption.

In [2] authors have proposed novel Attribute-Based Storage system which supports secure and efficient deduplication. It also explained drawback of standard Attribute-based encryption technique which does not support secure deduplication. The system works on hybrid cloud environment where private cloud is in charge of identical copies detection and public cloud opts for managing storage.

The system has two major advantages

1) Data Confidentiality is maintained while sharing data by specifying access policy.
2) The concept of data security is achieved here with high standard theory since other couldn't accomplish it under this theory.

In [3] author explained ABE (Attribute Based Encryption) technique used to reduce storage space and share data efficiently. In this system if attributes of particular user is matched then the person is given right to compute and decipher the enciphered data.

In [4] authors have introduced convergent encryption technique to secure data in process of deduplication of data. The data outsourced is converted to cipher text before performing deduplication. The authors have also introduced different privileges to the users.

In [5], authors have introduced (MLE) which provide secure deduplication. This scheme is best for large files as this needs schema perpetuation at servers. As large files needs better maintenance scheme suits it. This scheme supports both file-level and block-level deduplication.

In [6] authors have introduced updatable block-level deduplication which provides deduplication on encrypted data and easy updation of data. The issue in file level deduplication of effective updating of data is overcome here. Some challenges are overcome by MLE and others are effectively dealt by UBLD_e protocol. Dynamic Ownership management challenge is fulfilled here.

In [7] authors initiate idea to reduce the cost of updation of data. The existing MLE solution does not provide effective and secure updation of encrypted data to the user. The cost of updating single bit of data is quite high. So, the authors have introduced Updatable block-level message locked encryption technique which aims to reduce computation cost logarithm to file size. It has also introduced proof-of-ownership to users for access of files.

In [8], the author has introduced scheme which uses Symmetric Encryption algorithm, Hashing technique, Convergent encryption algorithm and token generation scheme to provide authorized duplication of data. Here the user data confidentiality and security is maintained. The data is protected both form passive and active attacks.

In [9] authors have introduced PoW (Proof-of-ownership) with data deduplication to support dynamic ownership management. This system support file-level, cross-user and block-level data deduplication. This scheme effectively carries out secure deduplication and maintains data confidentiality, consistency. It also reduces load of key management and storage space.

In [10] author has surveyed various methodologies and technologies for implementing data deduplication. They have also shown comparison of various technologies. The data confidentiality is compromised at different extent while performing data deduplication is depicted in the paper.

In [11] authors have introduced PoW (Proof-of-ownership) with data deduplication to support dynamic ownership management. This system support file-level, cross-user and block-level data deduplication. This scheme effectively carries out secure deduplication and maintains data confidentiality, consistency. It also reduces load of key management and storage space.

III. PROPOSED SYSTEM

The model consists of 3 things: A Cloud Service Provider (CSP), Users (U) and a Management Center (MC), as shown in Figure 1. The user makes request to MC for encrypting file uploaded by user for maintaining data confidentiality. Further the enciphered file is forwarded to the CSP [1].

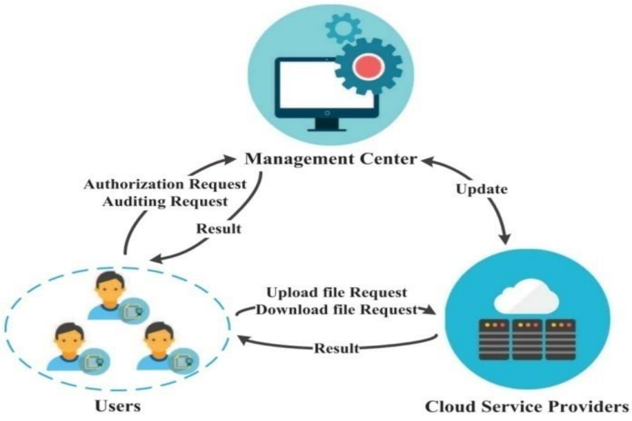

Figure 1 System model.

A. Entities of System Model

1) User

User belongs to different role groups having different role keys. Depending on control policies and role keys users can download and upload files from CSP. The creator of file is is special user and also unique.

2) CSP

Cloud service provider is responsible for data management, storage and verification. The file uploaded by the user is stored and managed by the CSP. CSP verify the user's identity and prevent unauthorized access.

3) MC

Management Center is trusted party and is responsible for role key management and user authorization.

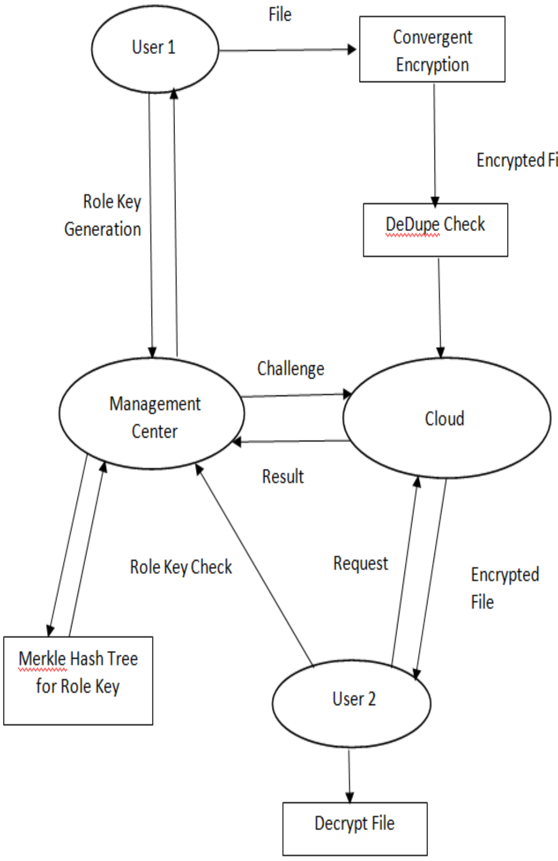

Figure 2. Flow of the system

B. Convergent Encryption

Convergent encryption algorithm is a symmetrical encryption algorithm. This algorithm encrypts the user file at the management center to maintain confidentiality of data while performing deduplication process [4], [17]. In this strong hash value is used to obtain convergent key from the original file. Hash value is is applied on the original file to obtain key. This key along with encryption algorithm is applied on file to get cipher text. Identical users who have identical files get identical hash values and identical keys to obtain identical cipher text.

Given original text file f, Encryption file Enc, ciphertext C, hash function h and convergent key k, we obtain

k = h (f)

C = Enck (f)

Convergent algorithm is used in SRRS to maintain user privacy and achieve secure deduplication of data in the cloud.

C. Merkle Tree

A hash tree or merkle tree is a tree structure in which each leaf node is a hash of a block of data and each non-leaf node is a hash of its children. This result in a single hash called the Merkle root. If every node has two children, the tree is called a binary hash tree. This merkle hash tree allows secure and efficient stoarge and handling of data. Since each node is generated from leaf node data cannot be altered or damaged thus maintaining confidentiality of data.

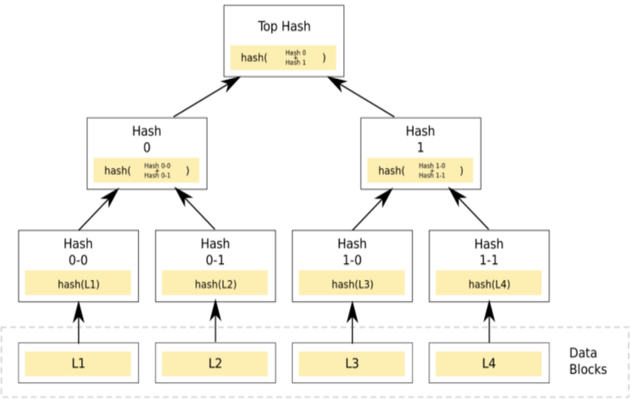

Fig. 2 Merkle Tree

D. Role Authorized Tree

In order to maintain role keys, role authorized tree is defined based on merkle hash tree, as shown in Figure 2. Management Center maintains this role authorized tree in order to manage authorized requests and confidentiality of data in the cloud. This tree has 2 nodes: leaf nodes and internal nodes. A leaf node contains data information and non-leaf or internal node does not have this information. Root node is considered under internal node.

E. Proof of ownership

In order to mainatin privacy of sensitive data proof of ownership is genearted. In case if any user supplies hash value to access the file PoW is implimented by CSP.

IV. CONCLUSION

In this paper our proposed system SRRS exploits convergent algorithm to maintain data privacy and role re-encryption algorithm to allow only authorized users to access data. This accomplishes authorized deduplication of data efficiently. Role Authorized tree maintains user role and role keys. Management center is introduced to maintain data confidentiality and reduce computation overhead.

In future better methods can be developed to achieve authorized deduplication and maintain security by preventing user's privacy information in the cloud computing.

V. REFERENCES

[1]. Jinbo xiong, Yuanyuan zhang, Shaohua tang, Ximengl liu and Zhiqiang Yao, "Secure encrypted data with authorized deduplication in cloud" IEEE Access, vol. 7, pp. 75090–75104, Jun.2019.

[2]. Hui cui , Robert H. deng, Yingjiu Li , Member and Guowei Wu, "Attribute-based storage supporting secure deduplication of encrypted data in cloud," IEEE transactions on big data, vol. 5, no. 3, July-September 2019.

[3]. Hua Ma1 , Ying Xie 1 , Jianfeng Wang2 , Guohua Tian1 , And Zhenhua Liu1, "Revocable attribute-based encryption scheme with efficient deduplication for e-health systems," Volume 7, 2019.

[4]. Jin li, Yan kit li, Xiaofeng chen, Patrick P.C. lee, and Wenjing lou," A hybrid cloud approach for secure authorized deduplication" IEEE transactions on Parallel and Distributed systems,. 2015.

[5]. Chen, R., Mu, Y., Yang, G., & Guo, F., " BL-MLE: Block-level message-locked encryption for secure large file deduplication", IEEE Transactions on Security, 2015.

[6]. Yongjun Zhao and Sherman S. M. Chow," Updatable block-level Message-locked encryption" Proc. IEEE Transaction on Dependable and secure computing, vol. xx, no. y, MAY 2019.

[7]. Maozhen Liu, Chao Yang, Qi Jiang, Xiaofeng Chen, Jianfeng Ma, Jian Ren, School of Cyber Engineering, Xidian University, Xi'an, Shaanxi, " Updatable block-level deduplication with dynamic ownership management on encrypted data".

[8]. Waghmare, V., & Kapse, S., "Authorized deduplication: An approach for secure cloud environment, 2016.

[9]. Hyungjune shin, Dongyoung koo, Youngjoo shin, and Junbeom hur," Privacy-preserving and updatable block-level data deduplication in cloud storage services" Proc. 2018 IEEE 11th International Conference on Cloud Computing.

[10]. Nipun Chhabra and Manju Bala,"A Comparative study of data deduplication strategies," in Proc. 2018 First International Conference on Secure

Cyber Computing and Communication (ICSCCC).

[11]. Shunrong Jiang , Tao Jiang and Liangmin Wang," Secure and Efficient cloud data deduplication with ownership management" Proc. IEEE Transaction, 2017.

[12]. Dapeng Wu, Hang Shi, Honggang Wang, Ruyan Wang, Hua Fang, "A feature-based learning system for Internet of Things applications," IEEE Internet Things J., vol. 6, no. 2, pp. 1928–1937, Apr. 2019.

[13]. J. Xiong, Y. Zhang, X. Li, M. Lin, Z. Yao, and G. Liu, "RSE-PoW: A role symmetric encryption pow scheme with authorized deduplication for multimedia data," Mobile Netw. Appl., vol. 23, no. 3, pp. 650–663, 2018.

[14]. W. Xia, H. Jiang, D. Feng, F. Douglis, P. Shilane, Y. Hua, M. Fu, Y. Zhang, and Y. Zhou, "A comprehensive study of the past, present, and future of data deduplication," Proc. IEEE, vol. 104, no. 9, pp. 1681–1710, Sep. 2016.

[15]. J. Li, C. Qin, P. P. C. Lee, and X. Zhang, "Information leakage in encrypted deduplication via frequency analysis," in Proc. 47th Annu. IEEE/IFIP Int. Conf. Dependable Syst. Netw., Jun. 2017, pp. 1–12.

[16]. J. R. Douceur, A. Adya, W. J. Bolosky, P. Simon, and M. Theimer, "Reclaiming space from duplicate files in a serverless distributed file system," in Proc. 22nd Int. Conf. Distrib. Comput. Syst., Jul. 2002, pp. 617–624

Cite this article as :

Prof. Milind B. Waghmare, Suhasini V. Padwekar , "Authorized Deduplication of Encrypted data in Cloud ", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 6 Issue 2, pp. 541-546, March-April 2020. Available at doi : https://doi.org/10.32628/CSEIT2062159
Journal URL : http://ijsrcseit.com/CSEIT2062159