International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2019 IJSRCSEIT | Volume 5 | Issue 2 | ISSN : 2456-3307
DOI : https://doi.org/10.32628/CSEIT1952223

Attribute Based Storage Supporting with Secure De-Duplication in Cloud

G. Akhila[1*], Ch. Jayanthi[2], G. Mounika[3], Ch. Likhita[4], T. Seshu Chakravarthy[5]
[1,2,3,4] Department of Computer Science and Engineering, VVIT, Guntur, Andhra Pradesh, India
[5]Assistant Professor, Department of Computer Science and Engineering, VVIT, Guntur, Andhra Pradesh, India
*Corresponding Author: akhilaganta123@gmail.com

ABSTRACT

Attribute based Encryption has been widely used in cloud computing for secure encryption of our files into the cloud. It controls the users who can access our data. It enables the users only with specific credentials to access our data. But this standard Attribute based Encryption does not support Deduplication of data, which helps in saving both the storage and network bandwidth of an organization. In this paper, we propose a system setting in a hybrid cloud, which will provide both features of secure encryption and data Deduplication.

Keywords - Cloud Computing, Deduplication, Hybrid cloud, ABE, CP-ABE.

I. INTRODUCTION

Data Deduplication is also called as single instance storage. It doesn't allow storing redundant copies of data. This technique ensures that only one unique instance of data is stored, and the redundant blocks are replaced with a pointer to unique data copy. Implementing this, Data De-duplication [1] in cloud can save a lot of money on storage costs to store the data and network bandwidth, while moving the data from one place to other. This technique helps organizations and cloud service providers greatly, because if you store less, you will pay less and hardware requirements to backup will also get reduced.

Cloud computing technology [2] is accessing data and programs over the internet, instead of Computer's Hard Drive. As cloud computing technology is growing day by day, outsourcing your data into cloud has become an attractive trend. At an organization level building your own infrastructure for storage and maintenance of data, is costly and tedious work. But by using cloud computing technology we can maintain our infrastructure to run the organization on pay as you go basis, which reduces cost, time of an organization.

To provide data confidentiality in the cloud, an encrypted file is saved into file and only the users with specific access credentials can be able to view or decrypt the file. But this type of file encryption does not support Deduplication. We cannot even upload the original file into the cloud as it is sensitive to security attacks. Since, the cloud is not fully trustworthy, implementation of Data De-duplication technique on the cloud generates many security concerns. In our proposed system, we implement Data deduplication by maintaining data integrity and data security by using a hybrid cloud [3].

The combination of both public and private cloud is called as hybrid cloud. In this, the private cloud will be responsible for Data Deduplication and the public cloud is responsible for storing the data. In this way we can provide security for the data and also implement the Deduplication techniques thus increasing the storage and network efficiency.

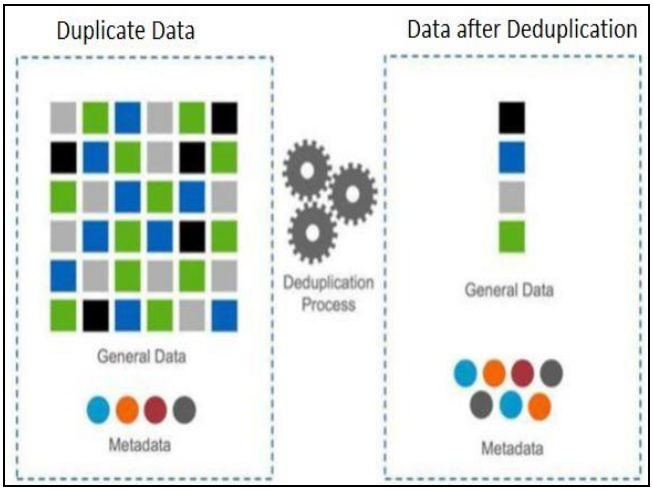

Fig 1: De-duplication

II. RELATED WORK

Data deduplication removes duplicate data copies in the storage. Generally, data deduplication can be divided into two basic approaches [4].

- Target-based data deduplication - It only focuses on escaping storing duplicate data but those duplicate data are still uploaded repeatedly. Therefore, it cannot advance the volume of transmissions.
- Source-based data deduplication - If the data have not been deposited, users need to upload the whole data, and the cloud storage server store them. Otherwise, users need to upload only the metadata; server creates the pointer, which stores to the first stored copy.

Deduplication is done in block level. In that, file is divided into blocks and that blocks are encrypted and create hash id by using different hashing techniques. Hash ids for blocks of file is created using algorithms like MD5, SHA variants etc. [5]. Hash ids are stored and check with the new hash ids when new file wants to put in cloud. Because of this technique, we may loss data security. It is vulnerable to brute force attacks and based on hash ids, we can re-generate the plaintext. Attackers have more scope to theft the sensitive data. We can't flexibly support to data access control.

We go for another encryption scheme i.e. convergent key encryption [6, 7] explains user encrypts data copy with convergent key which is derived from data. In addition, the user derives a tag for the data copy, such that the tag will be used to detect duplicates. We assume that tag correctness property holds [8]. Convergent Key generation is done with RSA algorithm [5]. Both convergent key and tag are independently derived and both encrypted data copy and tag are stored in server side.

ABE was proposed by Amit Sahai and Brent Waters [9]. In this Attribute based encryption, a user will encrypt their files for data security while putting file into cloud. Data encryption and decryption is done using different ABE schemes such as CP-ABE and KP-ABE [9, 10].

In KP-ABE, Access policies are associated with the key and cipher text is generated with the set of attributes, so that users who have the key will get the plaintext from cipher text. Main drawback of KP-ABE is the encrypted data cannot choose who can decrypt the file shared. The main drawback in KP-ABE is we do not know the users who are downloading files.

CP-ABE is reverse of KP-ABE. It is advanced version of KP-ABE. In CP-ABE, the cipher text is associated with access policies and the key is enclosed with set of attributes. Decryption is possible when attribute set follows the access policies. CP-ABE is more useful in terms of access control for cloud as it facilitates the data owner who can select the access policy to decide who is valid for downloading the cipher text.

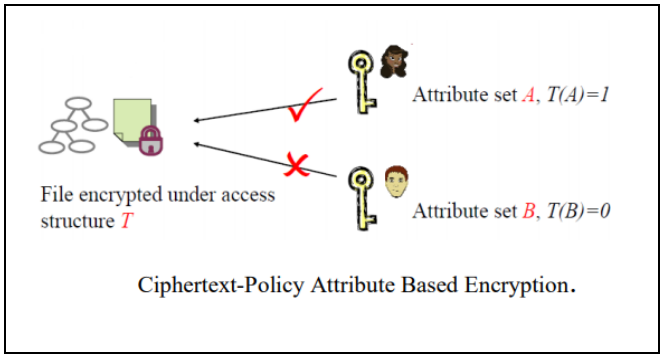

Fig 2: CP-ABE

Duplicate check is not possible in public cloud that's why hybrid set up is needed. K. Saritha et al., 2015 [12] Presented this approved duplicate check in hybrid cloud architecture. The hybrid cloud building suggests about both the public cloud and the private cloud. The private cloud plays a significant role in this system; moreover the security of this scheme is less, as the private cloud of this mechanism is not secure, unsanctioned access of the data resulting in fail in security. In order to offer more security, the private cloud is provided with multilevel authentication.

III. METHODOLOGY

In our proposed system, we use hybrid cloud architecture [3] in order to provide security and also deduplication. The tag generation and checking of these tags, providing access policies all these are done in private cloud, whereas public cloud is only used for storage of data. This model of deduplication is very useful for storage of data at an organization level, as it saves both the storage and network bandwidth, thus make it cost-effective.

1. Modules:

- Data provider: The person who uploads the data to store is called data owner. He is authorized by an administrator.
- User: The one who wants to download the files from cloud, and have the credentials to download the files are called users.
- Administrator: The person who authorizes the data providers and users, and provides credentials to download the data.

2. How our system works:

At any organization, when a data provider wants to upload the data into the cloud, He has to be authorized owner. Otherwise, he has to keep the request for administrator to authorize him. After validation user can upload a file into cloud. After the data provider provides the data encryption is done on that plaintext and a cipher text, tag, attributes are generated for that original plain text. This encrypted file is stored into the cloud. The data provider can also control who can view or download his data. He can provide access policies to the cloud, while uploading his file.

Only the users with these access policies will be able to view or download that specific file from the cloud. All this checking or providing credentials for a user is done by the administrator. When another data provider provides the data, first in the private cloud, after encryption it checks the tag of the file, with already existing tags of the files in the system. If it matches with any file, then it displays a message saying that the same file exists. Instead of saving the same file again, it writes the reference of the already existing file to the current file. In this way all this deduplication process is done at private cloud.

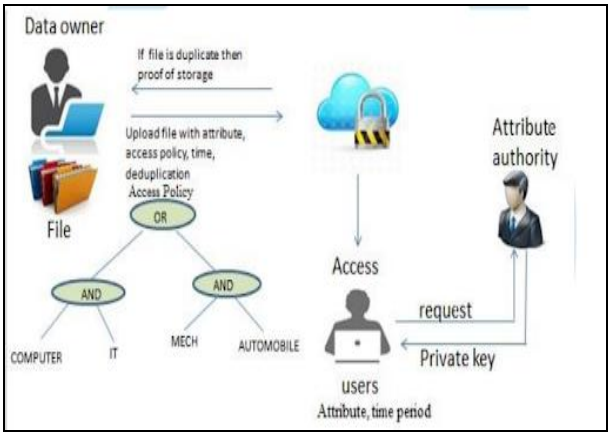

Fig 3: System Architecture

When a user wants to download any file from the cloud, he needs to have the specific credentials that have to be needed to download the file. If he doesn't have those credentials, he can request the administrator, to provide the credentials, the administrator then checks the user, and can accept or decline the request.

Ex: Let us take the organization as a hospital. In a hospital management system, we have a patient file, who was admitted to the hospital with 2 specific problems regarding neurotic and heart. The details of this patient are stored into cloud, by both cardiologist and neurologist regarding his disorders in 2 different files. If a receptionist wants to upload the same file regarding the same patient into cloud, instead of saving them again, the private cloud will provide references to the files that are already in the cloud. In this way deduplication is done. If a nurse want to view the details of the patient, he keeps request to the admin to provide the details, the admin then validate that nurse and checks whether he is belonging to the specific department that patient admitted or not, and then after checking all these details, the admin will provide credentials to that nurse regarding the patient.

IV. RESULTS AND DISCUSSION

In our system, we mainly concentrate on data de-duplication and we also work on data as a service to users in an organization. We provide interface to users to upload or download file and that organization gives this service to users through this interface only. Users or data providers are can't access the cloud. All transactions are done through this web interface only.

- Data de-duplication in cloud is done using hybrid cloud structure. We have 3 modules. Data Owner must be registered and validated by Admin.
- Data owner is logging in website for uploading the file into cloud.
- After login into website, we see the upload button to upload a file.
- If file is already exists, it displays error message of type filename same or Content same.
- If file is not uploaded before, it is uploaded successfully and before placed into cloud it is encrypted by admin and then placed in cloud by admin.
- When data user wants to download a file, he login and request for key to admin. Admin validates the user based on access policies that are enclosed to file and he get key if he is valid user. By using key, data user decrypts the data.
- Admin also know the downloaded persons details.

In previous work, data owner validates the user and send key to them but in this, Admin is responsible for user validation and key sending to valid user. So, downloading process is done quickly as compared to previous work [13]. We are doing this in hybrid cloud, so it is more secure.

V. CONCLUSION AND FUTURE SCOPE

Data deduplication in cloud is used to reduce the storage overhead created by the huge amount of duplicate data copies stored in the cloud. Here, Deduplication in private cloud is used to reduce size and improve the network band width, security and efficiency. Data security is protected by using

different permissions of users in the data duplication check. By using CP-ABE, we improve confidentiality even data is stored in untrusted server. This method is more secure against collision attacks. In future, we would like to generate a key and access policies to make data more secure.

VI. REFERENCES

[1]. D. Harnik, B. Pinkas, and A. Shulman-Peleg, "Side channels in cloud services: Deduplication in cloud storage", IEEE Security and Privacy, vol. 8, no. 6, pp. 40-47, 2010.

[2]. Bhairavi kesalkar, Dipali Bagade, Manjusha Barsagade, Namita Jakulwar, "Implementation of data deduplication using cloud computing" IJARIIT pp. 50 - 56.

[3]. Jin Li ,Yan Kit Li , Xiao feng Chen, Patrick P.C. Lee and Wenjing Lou, "A Hybrid Cloud Approach for Secure Authorized Deduplication" IEEE Transactions On Parallel And Distributed System Vol.26,No.5, May 2015.

[4]. Manreet kaur and Jaspreet Singh, "Data Deduplication Approach based on Hashing Techniques for Reducing Time Consumption over a cloud Network" in International journal of Computer Applications, vol. 142 - No.5, May 2016.

[5]. 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. ICDCS, 2002, pp. 617-624.

[6]. Jin Li, Xiaofeng Chen, Mingqiang Li, Jingwei Li, Patrick P.C. Lee, and Wenjing Lou, "Secure Deduplication with Efficient and Reliable Convergent Key Management" in IEEE transactions on Parallel and distributed Systems, vol. 25, No. 6, June 2014.

[7]. M. Bellare, S. Keelveedhi, and T. Ristenpart, "Message-Locked Encryption and Secure Deduplication," in Proc. IACR Cryptology ePrint Archive, 2012, pp. 296-3122012:631

[8]. Iuon-Chang and Po-Ching Chien, "Data deduplication Scheme for Cloud storage". IJ3C, Vol. 1, No. 2 (2012)

[9]. 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.

[10]. Sphurti Atram, and N. R. Borkar, "A review paper on Attribute-Based Encryption Scheme in Cloud Computing", IJCSMC, vol. 6, Issue. 5, May 2017, pg. 260-266.

[11]. 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, pp. 89=98.

[12]. Saritha K., and S. Subhasree "Analysis of hybrid cloud approach for private cloud in the de-duplication mechanism", Engineering and Technology (ICETECH), 2015 IEEE International Conference, 2015.

[13]. T. Seshu Chakravarthy, S. Rupa maheswari, A. L. V Priyanka, P. Asritha, and V. Divya keerthi, " Deduplication on Encrypted data in Cloud" in IRE journals, vol. 1, issue 9, 2018.

Cite this article as : G. Akhila, Ch. Jayanthi, G. Mounika, Ch. Likhita, T. Seshu Chakravarthy, "Attribute Based Storage Supporting with Secure De-Duplication in Cloud", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 5 Issue 2, pp. 786-790, March-April 2019. Available at doi : https://doi.org/10.32628/CSEIT1952223
Journal URL : http://ijsrcseit.com/CSEIT1952223