ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 3, Issue 2, pp: (301-305), Month: April - June 2015, Available at: www.researchpublish.com

Authenticated Deduplication System with Access Control and Security Measures

[1]Anu George, [2]Mr. Sandeep Hegde

[1]Dpt.Of Computer Science & Engg. Mangalore Intitute of technology & Engineering, Mangalore, Karnataka, India
[2]Assistant Professor, Dpt. Of Computer Science & Engg, Mangalore Institute of Technology and Engineering Mangalore, Karnataka, India

***Abstract*: Nowadays, cloud computing provides high amount of storage space and massive parallel computing at effective cost. It provides all kinds of services for the users. The main service offered by the cloud is nothing but storage facility. As cloud computing becomes popular, excessive amount of data being stored in the cloud. .One critical challenge of cloud storage services is the management of the ever-increasing volume of data. However, exponential growth of ever-increasing volume of data has raised many new challenges. De-duplication technique is specialized data compression technique which eliminates redundant data as well as improves storage and bandwidth utilization. Convergent encryption technique is proposed to enforce confidentiality during de-duplication, which encrypt data before outsourcing. To better protect data security, we present different privileges of user to address problem of authorized data de-duplication.**

***Keywords:* Deduplication, authorized duplicate check, confidentiality, hybrid cloud, covergent encryption.**

I. INTRODUCTION

Cloud computing will be the main information infrastructure in future. It provides all kinds of services for the users. The main service offered by the cloud is nothing but storage facility. One critical challenge of cloud storage services is the management of the ever-increasing volume of data. To make data management scalable and to reduce the increased amount of data in cloud ,deduplication has been a well-known technique. Data deduplication is a specialized data compression technique for eliminating duplicate copies of repeating data in storage. The technique is used to improve storage utilization and efficiency of cloud storage. deduplication allows to save storage space and minimize redundant data. In this model duplicate data is stored once and we keep pointers to the actual data. Traditional encryption requires different users to encrypt their data with their own keys. Thus, identical data copies of different users will lead to different cipher texts, making deduplication impossible. Deduplication can take place at either the file level or the block level. For file-level deduplication, it eliminates duplicate copies of the same file. Deduplication can also take place at the block level, which eliminates duplicate blocks of data that occur in non-identical files. Although data deduplication brings a lot of benefits, security and privacy concerns arise as users' sensitive data are susceptible to both insider and outsider attacks.

Convergent encryption is a widely used technique to combine the storage saving of de-duplication to enforce confidentiality. In convergent encryption, the data copy is encrypted under a key derived by hashing the data itself. This convergent key is used for encrypt and decrypt a data copy. After key generation and data encryption, users retain the keys and send the cipher text to the cloud. Since encryption is deterministic, identical data copies will generate the same convergent key and the same cipher text. This allows the cloud to perform de-duplication on the cipher texts. The cipher texts can only be decrypted by the corresponding data owners with their convergent keys. Differential authorization duplicate check is an authorized de-duplication technique where each user is issued a set of privileges during system initialization. This set of privileges specifies that which kind of users is allowed to perform duplicate check and access the files.

However, previous deduplication systems cannot support *differential authorization duplicate check*, which is important in many applications. In such an authorized deduplication system, each user is issued a set of privileges during system initialization .Each file uploaded to the cloud is also bounded by a set of privileges to specify which kind of users is allowed to perform the duplicate check and access the files. Before submitting his duplicate check request for some file, the user needs to take this file and his own privileges as inputs. The user is able to find a duplicate for this file if and only if there is a copy of this file and a matched privilege stored in cloud.

Traditional deduplication systems based on convergent encryption, although providing confidentiality to some extent, do not support the duplicate check with differential privileges. In other words, no differential privileges have been considered in the deduplication based on convergent encryption technique.

II. RELATED WORKS

cloud computing is now an emerging market. day by day application hosting on cloud increases rapidly causes huge data storage on cloud. due to this the main challenge faced by cloud service provider is the management of this ever-increasing bulk data.

S. Quinlan and S. Dorward. Venti[1] – In 2002, this paper presents an approach towards de-duplication called write-once policy of data. It provides efficient storage applications such as backup system i.e. logical backup, physical backup, and snapshot file systems.

J. R. Douceur, A. Adya, W. J. Bolosky, D. Simon, and M. Theimer[2] – This paper introduces convergent key technique. To enforce data confidentiality and making de-duplication feasible convergent encryption is proposed. By applying cryptographic hash function on data convergent key is generated. Using this key It encrypts/decrypts a data. Encrypted data is sent to the cloud and user preserves the key and sends the cipher text to the cloud. The encryption is deterministic operation. The key is derived from the data content, hence identical data copies will generate the same convergent key and using the same key same cipher text is generated.

Pinkas, and A. Shulman-Peleg[3] – To prevent unauthorized access, a secure proof of ownership protocol is also needed to provide the proof that the user indeed owns the same file when a duplicate is found. After the proof, subsequent users with the same file will be provided a pointer from the server without needing to upload the same file. A user can download the encrypted file with the pointer from the server, which can only be decrypted by the corresponding data owners with their convergent keys.

M. Bellare, S. Keelveedhi, and T. Ristenpart[4] –Message-locked encryption and secure de-duplication: In this they formalize a new cryptographic primitive that they call Message-Locked Encryption (MLE), where the key under which encryption and decryption are performed is itself derived from the message. MLE provides a way to achieve secure de-duplication, a goal currently targeted by numerous cloud storage providers.

Weak leakage-resilient client-side de-duplication of encrypted data in cloud storage by Xu et al.[5]- also addressed the problem and showed a secure convergent encryption for efficient encryption.The proposed technique only focuses on encryption and file level de-duplication. The issue of key-management and block-level de-duplication is not considered.

D. Ferraiolo and R. Kuhn. [6] – Role-based access controls: In this they represent limitation of Mandatory Access Controls (MAC) technique. This is required for high level security like multilevel secure military applications.

Architecture for secure cloud computing - Bugiel et al. [8] – It provided an architecture consisting of twin clouds for securely outsourcing of user private data and arbitrary computations to an untrusted commodity cloud.

Zhang et [9] al also presented the hybrid cloud techniques to support privacy-aware data-intensive computing. We consider addressing the authorized privileged de-duplication problem over data in public cloud. The security model of our systems is similar to those related work, where the private cloud is assume to be honest but curious.

S. Halevi, D. Harnik, B. Pinkas, and A. Shulman-[10] – Proposes of POW (proof of ownership) technique is that a user can efficiently prove to the cloud storage server that he/she owns a file without uploading the file itself. It also proposes the Merkle-Hash Tree to enable client-side de-duplication, which include the bounded leakage setting. The proposed scheme is focusing only on the data ownership and not on the data privacy.

In S. Ossowski and P. Lecca: [11] extended proofs of ownership mechanism for encrypted files. These papers do not address how to minimize the key management overhead.

Pietro and Sorniotti [14] proposed another efficient PoW scheme by choosing the projection of a file onto some randomly selected bit-positions as the file proof But this project do not deal with data privacy.

III. SYSTEM ARCHITECTURE

In this paper, we address the problem of privacy preserving deduplication in cloud computing and propose a new deduplication system that includes the public cloud and the private cloud and also the hybrid cloud which is a combination of the both public cloud and private cloud. In general if we used the public cloud we can't provide the security to our private data and hence our private data will be loss. So that we have to provide the security to our data for that we make a use of private cloud also. When we use a private clouds the greater security can be provided. In this system we also provides the data deduplication. which is used to avoid the duplicate copies of data. User can upload and download the files from public cloud but private cloud provides the security for that data. that means only the authorized person can upload and download the files from the public cloud.for that user generates the key and stored that key onto the private cloud.at the time of downloading user request to the private cloud for key and then access that Particular file.

In our architecture there are three modules.

[1] user

[2] public cloud

[3] private cloud. etc

First if the user want to upload the files on the public cloud then user first encrypt that file with the convergent key and then sends it to the public cloud at the same time user also generates the key for that file and sends that key to the private cloud for the purpose of security.In the public cloud we use one algorithm for deduplication. Which is used to avoid the duplicate copies of files which is entered in the public cloud. Hence it also minimizes the bandwidth. that means we requires the less storage space for storing the files on the public cloud.In the public cloud any person that means the unauthorized person can also access or store the data so we can conclude that in the public cloud the security is not provided. In general for providing more security user can use the private cloud instead of using the public cloud.User generates the key at the time of uploading file and store it to the private cloud.When user wants to downloads the file that he/she upload,.he/she sends the request to the public cloud. Public cloud provides the list of files that are uploads the many user of the public cloud because there is no security is provided in the public cloud.When user selects one of the file from the list of files then private cloud sends a message like enter the key!. User has to enter the key that he generated for that file. When user enter the key the private cloud checks the key for that file and if the key is correct that means user is valid then private cloud give access to that user to download that file successufully. then user downloads the file from the public cloud and decrypt that file by using the same convergent key which is used at the time of encrypt that file.in this way user can make a use of the architecture.

S-CSP: The purpose of this entity to work as a data storage service in public cloud. The S-CSP eliminate the duplicate data using deduplication and keep the unique data as it is. S-SCP entity is used to reduce the storage cost. S-CSP has abundant storage capacity and computational power.

Data User: A user is an entity that want to access the data or files from S-SCP.User generate the key and store that key in private cloud. In storage system supporting deduplication, The user only upload unique data but do not upload any duplicate data to save the upload bandwidth, which may be owned by the same user or different users. Each file is protected by convergent encryption key and can access by only authorized person. In our system user must need to register in private cloud for storing token with respective file which are store on public cloud. When he want to access that file he access respective token from private cloud and then access his files from public cloud.

Private Cloud: In general for providing more security user can use the private cloud instead of public cloud.User store the generated key in private cloud.At the time of downloading system ask the key to download the file. User can not store the secrete key internally. for providing proper protection to key we use private cloud. Private cloud only store the

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 3, Issue 2, pp: (301-305), Month: April - June 2015, Available at: www.researchpublish.com

convergent key with respective file. When user want to access the key he first check authority of user then an then provide key.

Public Cloud: Public cloud entity is used for the storage purpose .User upload the files in public cloud. Public cloud is similar as S-CSP.When the user want to download the files from public cloud, it will be ask the key which is generated or stored in private cloud. When the users key is match with files key at that time user can download the file, without key user can not access the file. Only authorized user can access the file. In public cloud all files are stored in encrypted format. If any chance unauthorized person hack our file, but without the secrete or convergent key he doesn't access original file. On public cloud there are lots of files are store each user access its respective file if its token matches with S-CSP server token.

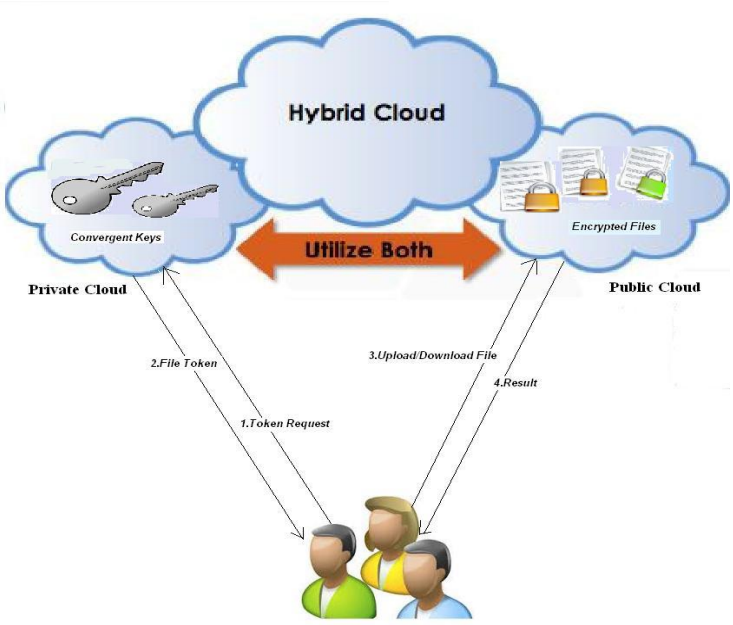

Fig.1 Architecture of Authorized Deduplication

Operations performed on Hybrid Cloud:

File Uploading : When user want to upload the file to the public cloud then user first encrypt the file which is to be upload by make a use of the symmetric key, and send it to the Public cloud. At the same time user generates the key for that file and sends it to the private cloud. in this way user can upload the file in to the public cloud.

File Downloading: When user wants to download the file that he/she has upload on the public cloud. he/she make a request to the public cloud. then public cloud provide a list of files that many users are upload on it. Among that user select one of the file from the list of files and enter the download option at that time private cloud sends a message that enter the key for the file generated by the user. then user enters the key for the file that he/she is generated. then private cloud checks the key for that file and if the key is correct that means the user is valid. only then the user can download the file from the public cloud otherwise user can't download the file. When user download the file from the public cloud it is in the encrypted format then user decrypt that file by using the same symmetric key.

IV. CONCLUSIONS

In this paper, the idea of authorized data deduplication was proposed to protect the data security by including differential authority of users in the duplicate check. In public cloud our data are securely store in encrypted format, and also in private cloud our key is store with respective file. There is no need to user remember the key. So without key anyone can not access our file or data from public cloud.This paper will provide more efficiency and security in cloud computing using authorized deduplication check and hierarchical access control method. It will improve the storage efficiency and perfomance of cloud storage.

REFERENCES

[1] S. Quinlan and S. Dorward. Venti:"a new approach to archival storage", In Proc. USENIX FAST, Jan 2002

[2] J. R. Douceur, A. Adya, W. J. Bolosky, D. Simon, and M. Theimer. "eclaiming space from duplicate files in a server less distributed file system.", In ICDCS, pages 617–624, 2002. S. Halevi, D. Harnik, B.

[3] Pinkas, and A. Shulman-Peleg. "Proofs of ownership in remote storage systems." In Y. Chen, G. Danezis, and V. Shmatikov, editors, ACM Conference on Computer and Communications Security, pages 491–500. ACM, 2011

[4] M. Bellare, S. Keelveedhi, and T. Ristenpart. "Dupless: Serveraided encryption for deduplicated storage." In USENIX Security Symposium, 2013.

[5] J. Xu, E.-C. Chang, and J. Zhou. "Weak leakage-resilient client-side de-duplication of encrypted data in cloud storage." , In ASIACCS,

[6] D. Ferraiolo and R. Kuhn. "Role-based access controls." In 15th NIST-NCSC National Computer Security Conf., 1992.

[7] K. Zhang, X. Zhou, Y. Chen, X.Wang, and Y. Ruan. "Sedic: privacyaware data intensive computing on hybrid clouds.", In Proceedings of the 18th ACM conference on Computer and communications security, CCS'11, pages 515–526, New York, NY, USA, 2011. ACM.

[8] S. Bugiel, S. Nurnberger, A. Sadeghi, and T. Schneider. "Twin clouds: An architecture for secure cloud computing." In Workshopon Cryptography and Security in Clouds (WCSC 2011), 2011.

[9] K. Zhang, X. Zhou, Y. Chen, X.Wang, and Y. Ruan. "Sedic: privacyaware data intensive computing on hybrid clouds.", In Proceedings of the 18th ACM conference on Computer and communications security, CCS'11, pages 515–526, New York, NY, USA, 2011. ACM.

[10] S. Halevi, D. Harnik, B. Pinkas, and A. Shulman-Peleg. "Proofs of ownership in remote storage systems." In Y. Chen, G. Danezis, and V. Shmatikov, editors, ACM Conference on Computer and Communications Security, pages 491–500. ACM, 2011.

[11] W. K. Ng, Y. Wen, and H. Zhu. "Private data de-duplication protocols in cloud storage." In S. Ossowski and P. Lecca, editors, Proceedings of the 27th Annual ACM Symposium on Applied Computing, pages 441–446. ACM, 2012.

[12] R. D. Pietro and A. Sorniotti. "Boosting efficiency and security in proof of ownership for de-duplication." In H. Y. Youm and Y. Won, editors, ACM Symposium on Information, Computer and Communications Security, pages 81–82. ACM, 2012.

Author's Profile:

Anu George: completed the bachelors degree in Computer Science &Engineering from anna university Chennai and presently pursuing Masters in Engineering in Computer Science & Engineering at Mangalore Institute of Technology, Mangalore.

Sandeep Hegde: completed bachelors and masters degree in Computer Science and Engineering. Currently working as a assistant professor in Manglore institute ofTechnology and Engineerig,Manglore. he has 3 Years Of industrial experience as a software engineer in Tata Consultancy Service.He has Presented a paper in National conference.