A Modern authorized de-duplication method on hybrid cloud

Vinayak Mehta

Master of Computer Application
VESIT, Chembur, Mumbai
University of Mumbai
vinayak.mehta508@gmail.com

Abstract: *Data compression techniques are widely used in order to avoid repeated data's which occupy large amount of space and bandwidth. Data duplication plays a vital role on data compression by effectively eradicate the duplicate copies of data. A care should be taken on data confidentially and privacy while doing the data de-duplication, to achieve this convergent encryption technique is proposed by encrypting the data before outsourcing. Initially for secure data transmission, problem of authorized data de-duplication is an issue. Considering the traditional methods data duplication must be checked within the data itself. We implement the new de-duplication constructions by supporting duplication check on data with secure authorization in hybrid cloud architecture. Our proposed system implies level of privacy and secure of sensitive data by implementing a new de-duplication security model. To ensure the efficiency of our proposed system various duplication checking and testing experiment are conducted using our prototype. As a result the new system increases the authorization and secure level in maximum when compared to the tradition systems.*

Key Words: **Hybrid Cloud, privacy, sensitivity, authorized duplicate check, De-duplication**

1. INTRODUCTION:

According to today's globalization cloud computing allows various facilities for the users in several aspects such as Naas, Paas, Saas and Iaas. In the current world cloud providers facilities storage space and vast parallel computing sources in a minimum cost. Cloud computing enables various security aspects in order to provide the client users for accessing the data and rights on the stored data. The important issue on the cloud storage services is managing of large amount of data. A scalable data management is possible by a well-known data computing application such de-duplication. The reason for this massive effectiveness is problem of repeated data occupies the large amount of storage space which affects the overall cloud computing process of a network. On discussing about de-duplication [1] it is an improvised data compression technique avoids repeated data to be stored. It not only occupies the spaces, by processing the overall data during transmission it increase the bandwidth as well as overload package. Having content with same data, de-duplication enables the original data by keeping only one physical copy instead of multiple ones by referring the repeated data to that copy. De-duplication can be applicable to place on both levels such as file level or block level. On discussing about file level it checks the repeated data within that file. To the other aspect on block level it avoids data copies which are present at the non-identical files. Data de-duplication provides not only space but also security and privacy on the user's sensitive data which was the serious issue on the client aspect. On de-duplication data privacy must be secured from the malicious attackers within as well as outside the network. To make this privilege encryption concept is evolved there are various traditional encryption techniques are followed to ensure the data confidentiality in a network. Because most of the existing methods requires the data owner to encrypt their data with own keys. But the problem is unexpectedly the data from different owners enables multiple cipher text that makes the de-duplication impractical. To overcome this problem convergent encryption [2] is proposed that makes possible of data de-duplication along with ensuring the data confidentiality. The actual methodology behind this approach is the encryption/decryption on the repeated data by applying cryptographic hash value on that content. The process happens in such as a manner that retains the keys and sends the encrypted data to the cloud. By doing this the same data copies will outcomes with similar convergent keys as well as the cipher text. The next this is to secure those data from the unauthorized users to achieve this separate protocol[3] is required to owns the file if there any duplicated data's are found. By the proof the applicable file is provided from the server instead of uploading the same one. That file can be downloaded by the authorized user and can be decrypted only with the convergent keys otherwise the data cannot decrypt which proves the data security to the owner. By doing this methodology it proven that cloud makes the de-duplication and convergent prevent the sensitive data from the unauthorized users.

1.1. Related Works:

The emerge of cloud computing and data security In earlier existing de-duplication method does not allows multiple authorization to check the duplication. On those systems each individuals were given a privilege during the system initialization. This attracts the research circle to cloud de-duplication according to Yuan et al.[4] his de-duplication on cloud storage system minimize the storage size by tagging integrity checks. By that he enables data de-duplication and provides data privacy. Bellare et al.[5] he states data protection can be possible by transforming the predictable message into unpredictable messages. On that stage he evolves the concept of third party a key server who generates the file tag for duplicate check. Stanek et al.[6] shows a novel encryption system which provide different security for popular and unpopular data. He applied a normal encrypted method for unpopular data and for popular data he performs two-layered encryption scheme with stronger security while supporting de-duplication by this way proved an efficient result on outsourced data. By Li et al.[7] he achieved block-level de-duplication on distributing the keys by multiple servers after encrypting the files. Xu et al.[8] gives the convergent encryption as an efficient encryption, without considering issues one key-management and block-level de-duplication. Some of other convergent encryption variants for secure de-duplication (e.g., [9], [10], 11],[12]) are also discussed. As we know the bitcasa a commercial cloud provider use this convergent encryption for reliable process.

1.2.Contribution:

In this paper aimed to give effective cloud computing process by implementing data de-duplication as an effective data compression technique. For that we consider hybrid cloud architecture which consists of private cloud and public cloud. But not like the traditional data de-duplication system by making the private cloud as a proxy to allow data owner/users to securely perform duplicate check with differential privileges. In this proposed architecture the data in the public cloud is outsourced which are managed in the private cloud. The duplication check is implemented which allows only the user to do the duplicate check for files marked with the corresponding privileges.

2.PROPOSED SYSYTEM

We propose an efficient de-duplication technique with new approach of authorization technique. In our system a hybrid cloud architecture for solving the problem. The key method is, the access permission is by providing private key will not directly provided to the user that is process by private cloud server separately. The private is unique and cannot be shared by the other users. In order to get the token the user first sends the request to service provider because to check the duplication data in the file the user must have the token access. At the same time the cloud provider also check the user token identity whether is an authorized or not. After the confirmation only the process will be continued. In accordance with public cloud the user check duplication of data before the file to be uploaded. The result proves the data efficiency either the file to be uploaded or re-considered. On the construction of our authorized de-duplication method has overcome the several problems such as sharing of private keys between the users to compute the file makes the authorized de-duplication system limited. The next things are it cannot prevent the privilege of sharing the private key between the users. Another important obstacle is brute-force attacks that can recover files falling into a known set. As a result the de-duplication system cannot provide privacy of predictable files. Because, proposed convergent encryption system must protect only the semantic security of unpredictable files. Thus the above discussed drawbacks are efficiently overcome successfully and make our new approach prominent.

3.IMPLEMATATION:

We set the hybrid cloud architecture in an enterprise network with multiple clients such as employees of an organization for whom the data de-duplication method will be in an effective manner. Because there a vast among of data must use to store in the cloud and maximum possibility of repeated data can be possible. At the same time the data de-duplication can set frequently data backup and disaster recovery applications in minimum memory space on the cloud. Those methods are more applicable for file backup and synchronization applications than richer storage abstractions. Our proposed system has entities such as user private cloud and public cloud.

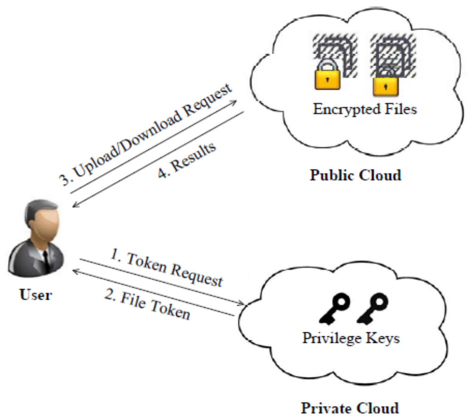

Fig 1: Architecture of Authorized De-duplication

We can say in our model both private cloud and public cloud are honest but curious, but buy our proposed protocol we can find the secret information as possible by their possessions in the network. By the enabled privilege the user tries to access the data within or out of the scopes. We make or technology with more efficient that is if all the file are sensitive and these are to be protect fully from both public and private cloud than we use two aspects such that aimed to extract more secret information as much as possible on both public cloud and private cloud in a entire network. The next thing is on internal side collection more information on the file from the public cloud and from outside scope duplicate-check token information from the private cloud is done.

3.1. Symmetric encryption:

The symmetric key uses a single secret key for encryption as well as decryption of information. That is the initial stage has perform the key generation and next the encryption is done with key along with message have the encrypted cipher text. The final thing is decryption of original data by using the same secret key.

3.2. Convergent encryption:

Convergent encryption makes de-duplication prominent it retains the data privacy in a maximum level. The working methodology is ever individual has the convergent key from the original data copy and encrypts those data copy with the same convergent key. By doing this the user has a tag based on the tag only duplicate data will be identified. Because if two or more data resembles same, than their tags also will be same in such a manner the duplications are removed. Such that which are identical will stored the data in the cloud and those tag cannot be used to deduce the convergent key which result in breaking the data privacy.

Function call for token generation and file uploading process:

- `FileTag(File)` - It computes SHA-1 hash of the File as File Tag;
- `TokenReq(Tag, UserID)` - It requests the Private Server for File Token generation with the File Tag and User ID;
- `DupCheckReq(Token)` - It requests the Storage Server for Duplicate Check of the File by sending the file token received from private server;
- `ShareTokenReq(Tag, {Priv.})` - It requests the Private Server to generate the Share File Token with the File Tag and Target Sharing Privilege Set;
- `FileEncrypt(File)` - It encrypts the File with Convergent Encryption using 256-bit AES algorithm in cipher block chaining (CBC) mode, where the convergent key is from SHA-256 Hashing of the file; and
- `FileUploadReq(FileID, File, Token)` - It uploads the File Data to the Storage Server if the file is Unique and updates the File Token stored.

Key maintenance of the private server:

- `TokenGen(Tag, UserID)` - It loads the associated privilege keys of the user and generate the token with HMAC-SHA-1 algorithm; and
- `ShareTokenGen(Tag, {Priv.})` - It generates the share token with the corresponding privilege keys of the sharing privilege set with HMAC-SHA-1 algorithm.
- `DupCheck(Token)` - It searches the File to Token Map for Duplicate; and
- `FileStore(FileID, File, Token)` - It stores the File on Disk and updates the Mapping.

4. RESULT & DISCUSSION:

4.1. Data user:

In a network the client user is an entity who wants to store the data on the cloud and retrieve the data when he wants without affecting the data privacy. In a storage a user can able to upload the file in the cloud but he could not able to check the uploaded data is unique are withholds the duplicated data on the file. Because it may require large space as well as bandwidth while processing. But in our system each user has the privilege to set up the system that is file is protected with the convergent encryption key which enables authorized de-duplication.

4.2. Hybrid Cloud:

Not as the traditional method our proposed system has both private and public cloud they can be cluster with trusted virtualized cryptographic co-processors. The result is third party who provide the require hardware based safety aspects in order to implement a remote execution environment which is trusted by the users.

4.3. Authorized duplicate check:

On proving the data de-duplication each user have convergent key along with tag that enables while uploading if multiple data copy are formed that the same tags are shown which shows the data duplication. If the data tags are unique than the uploaded the having data's are also unique. In same aspect the security is ensured by the convergent encryption technique which provides unique key with certain privilege to the user that was discussed earlier.

4.4. File Uploading :

If the data owner wants to upload his files than he can share the file with his privilege. The owner of data interact with public cloud that done duplication check before uploading the file. It can be done by private key and process send by client-server method on which the data originality is confirmed then it is encrypted before uploading to ensure the data privacy on the hybrid cloud.

4.5. File retrieving:

The user downloads the file in the same manner with the private key privileges to decrypt the file and view the original data from the cloud. In rather if it fails to download the file by providing the wrong key that the user is not the original user which achieves the primary goal data security in a cloud.

4.6. Data Confidentiality:

In detail that the unauthorized users who not having the proper privilege or file from private cloud server is prevented from access to the underlying plaintext stored. Which means retrieve and recover the files that do not belong to them. That proven that convergent encryption based data confidentiality is the higher level confidentiality as compared to the traditional approaches.

5. CONCLUSION:

In this paper the authorized data de-duplication is focused on the real time issue of duplicated data and privacy breach in a cloud. Our proposed system protects the data and proved the absence of repeated data in effective manner by an approach convergent encryption technique in hybrid cloud architecture. Our security schemes prove it capacity in both inside and outside attacks which was specified in the proposed security model. In addition our proposed authorized duplicate check scheme undergoes various tested experiment under our prototype to guarantee its effectives by stating our authorized duplicate check scheme achieves minimum overhead compared
to convergent encryption and network transfer which are considered to the traditional methods.

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. Reclaiming space from duplicate files in a serverless distributed file system. In ICDCS, pages 617–624, 2002.
[3] J. Yuan and S. Yu. Secure and constant cost public cloud storage auditing with deduplication. IACR Cryptology ePrint Archive, 2013:149, 2013.
[4] M. Bellare, S. Keelveedhi, and T. Ristenpart. Dupless: Serveraided encryption for deduplicated storage. In USENIX Security Symposium, 2013
[5] J. Stanek, A. Sorniotti, E. Androulaki, and L. Kencl. A secure data de duplication scheme for cloud storage. In Technical Report, 2013.
[6] J. Li, X. Chen, M. Li, J. Li, P. Lee, andW. Lou. Secure de duplication with efficient and reliable convergent key management. In IEEE Transactions on Parallel and Distributed Systems, 2013.
[7] J. X u, E.-C. Chang, and J. Zhou. Weak leakage-resilient client-side de duplication of encrypted data in cloud storage. In ASIACCS, pages 195–206, 2013.
[8] P. Anderson and L. Zhang. Fast and secure laptop backups with encrypted de-duplication. In Proc. of USENIX LISA, 2010.
[9] A. Rahumed, H. C. H. Chen, Y. Tang, P. P. C. Lee, and J. C. S. Lui. A secure cloud backup system with assured deletion and version control. In 3rd International Workshop on Security in Cloud Computing, 2011
[10] M. W. Storer, K. Greenan, D. D. E. Long, and E. L. Miller. Secure data de duplication. In Proc. of StorageSS, 2008.