International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2018 IJSRCSEIT | Volume 4 | Issue 2 | ISSN : 2456-3307

Compreive Anasysis On Secure Data Deduplication With Dynamic Ownership Management In Cloud Data Storage

Chavva. Ravi Kishore Reddy*[1], Kotapati Naga Phaneendra*[2]
[1*]ravikishore63@gmail.com, phanek22@gmail.com[2]

ABSTRACT

In cloud storage services, deduplication innovation were acquainted with lessen the space and transmission capacity necessities of services by taking out repetitive information and putting away just a solitary duplicate of them. Deduplication is best when various clients transfer similar information to the cloud storage, yet it raises issues identifying with security and ownership. Evidence of-ownership plans permit any proprietor of similar information to demonstrate to the cloud storage server that he claims the information in an unexpected way. As of late, a few Deduplication plans are proposed to tackle this issue by enabling every proprietor to have a similar encryption key for similar information. Notwithstanding, the greater part of the plans experience the ill effects of security issues, since they don't consider the dynamic changes in the ownership of the information that happen much of the time in a reasonable cloud storage benefit. In this paper, we propose a server-side Deduplication conspires for encoded information. It enables the cloud server to control access to outsourced information notwithstanding when the ownership changes powerfully by giving focalized encryption and secure ownership gather key circulation. This averts information spillage not exclusively to renounced clients despite the fact that they beforehand possessed that information, yet additionally to any genuine inquisitive cloud storage server. Moreover, the proposed conspire ensures information trustworthiness against any label irregularity assault. Subsequently, security is improved in the proposed plot. The productivity investigation comes about show that the proposed plot is nearly as effective as the past plans, while the extra computational overhead is insignificant.

Keywords : Deduplication, Cloud Storage, Encryption, Proof-of-Ownership

I. INTRODUCTION

Cloud Computing is a widespread term used in today's world. It delivers infinite space for storage, readiness, user-friendliness from anywhere, anytime to entities. Now-a-day's number of users and their data in the cloud is continuously growing with higher memory space and upload bandwidth. Data de-duplication used in cloud storage providers to resolve these overheads. Deduplication is a process of removing multiple copies of same data, to reduce the storage space and save bandwidth. But when same data outsourced by users to cloud storage some challenges are arises on data ownership and security for sensitive data.

Today's cloud storage services like Drop box and Google Drive etc. use a de-duplication scheme to save the network bandwidth and the storage cost. As data owners worried about their private data, they may encrypt their data before uploading in order to keep data privacy from illegal outside adversaries, as well as from the cloud service provider. As concern with authorized access and security, there are many encryption schemes proposed. De-duplication scheme takes benefit of data similarity to find the same data and scale down the storage space. In

CSEIT1833613 | Received : 01 April 2018 | Accepted : 10 April 2018 | March-April-2018 [(4) 2 : 57-63]

contrast, encryption algorithms randomized the encrypted files to make cipher-text same from theoretically random data. Encryption of the same data by dissimilar users with different encryption keys results in different cipher texts, which makes it hard for the cloud server to decide whether the plain data are the same and de-duplicate them. Hence, traditional encryption makes de-duplication impossible for above reasons. The simplest implementation of traditional encryption can define as follows: Consider users A and B, encrypts the same file M under their secret keys SKA and SKB and stores corresponding cipher-text CA and CB. Then, further problems arise: First, how can the cloud server sense that the underlying file M is similar, and second is even if it can notice this, how can it allow both users to recover the stored data, based on their distinct secret keys? One simple way out is to let on each client to encrypt the file with the public key of the cloud storage server. Then, the server is capable to de-duplicate the identified data by decrypting it with its private key pair. Still, this solution grants access to the cloud storage server to get the outsourced plain data, which may break up the privacy of the data if the cloud server cannot be fully trusted. Convergent encryption plays the vital role in data Deduplication and overcomes the drawback which discussed above. A convergent encryption algorithm works as follows: Firstly, it takes an input file and encrypts them with its hash value as an encryption key. Then, the cipher text is given to the cloud server and user keeps the encryption key. As convergent encryption is deterministic, every time similar files encrypted into similar cipher-text irrespective of who encrypts them Hence, the cloud server can do de-duplication over the generated cipher text. Then all data owners can download the cipher text and decrypt it later as they have the same encryption key for the file. But convergent encryption has security weakness concern with tag consistency and ownership revocation. This paper formalizes a scheme to solve the challenge of ownership changes dynamically in the cloud system.

II. LITERATURE SURVEY

In cloud computing, there have been many of the schemes, proposed for data Deduplication over encrypted and unencrypted data of cloud storage. We are going to discuss about the data Deduplication schemes over encrypted data and how it has been developed and improved further into Convergent Encryption (CE), Leakage-Resilient (LR) Deduplication scheme, Randomized Convergent Encryption (RCE) and Dynamic Ownership Management Scheme.

Convergent Encryption (CE): LI [1] In order to keep data privacy against inside cloud server as well as outside challengers, users may want their data encrypted. However, conventional encryption under different users' keys makes cross-user de-duplication impossible, since the cloud server would always see different cipher texts, even if the data are the same, regardless of whether the encryption algorithm is deterministic. Douceur introduces Convergent Encryption, which is the promising solution to this problem. In CE, a data owner derives an encryption key over data by using cryptographic hash function. Then computes the cipher text using block cipher over data along with their encryption key. CE deletes data and keeps only encryption key after uploading cipher text to the cloud storage. Since encryption is deterministic, on receipt of same file CE generates same cipher text for it and the server does not store the file but instead updates meta-data to indicate it has an additional owner.

Advantages: Provides promising solution over conventional encryption and preserves data privacy.

Disadvantages: Convergent Encryption suffers from some security issues i.e. tag consistency problem. It means that integrity and security of data has been compromised due to the lack of proof of ownership process and dynamic ownership management.

Ramp Secret Sharing Scheme (RSSS) LI[2] formalizes a convergent key management scheme i.e. Dekey which is efficient and reliable for secure deduplication. Dekey set de-duplication between convergent keys and distributes those keys across multiple key servers while preserving the semantic security of convergent keys and privacy of outsourced data. Dekey is implemented using the Ramp secret sharing scheme. Dekey uses RSSS to collect convergent keys. Its idea is to permit deduplication in convergent keys and distribute the convergent keys over various KM-CSPs. Instead of encrypting the convergent keys on a per-user basis, Dekey builds secret shares on the original convergent keys (that are in plain) and assigns the shares over various KMCSPs. If many users share the identical block, they can access the same corresponding convergent key. This significantly decrease the storage overhead for convergent keys. In addition, this method provides fault tolerance and allows the convergent keys to remains accessible even if any subset of KM-CSPs fails.

Advantages: Provides reliable, efficient and fault tolerance convergent key mechanism for secure de-duplication.

Disadvantages: This scheme does not support dynamic ownership management issue in secure de-duplication.

Authorized De-duplication Hybrid Cloud LI[3] proposes an authorized de-duplication scheme where differential privileges of users, as well as the data, are considered in the de-duplication procedure in a hybrid cloud environment. He presented several new de-duplication constructions supporting authorized duplicate check in hybrid cloud architecture, in which the duplicate check tokens of files are generated by the private cloud server with private keys. The figure shows the architecture of authorized de-duplication.

Advantages: This scheme provides authorized de-duplication over hybrid cloud for users who have different privileges.

Disadvantages: Data leakage.

Cloud Computing and Emerging IT Platforms this paper, author characterize Cloud computing and give the structural planning to making Clouds with business sector arranged.Resource allocation by utilizing advancements, for example, Virtual Machines (VMs). Authors additionally give bits of knowledge on market-based resource administration systems that incorporate both client driven service management and computational risk administration to manage Service Level Agreement (SLA) - arranged resource distribution. What's more, authors uncover our initial musings on interconnecting Clouds for progressively making worldwide Cloud trades and markets. At that point, we display some illustrative Cloud stages, particularly those created in commercial enterprises alongside our present work towards acknowledging market-situated resource portion of Clouds as acknowledged in Aneka venture Cloud innovation. Besides, author highlight the distinction between High Performance Computing (HPC) workload furthermore, Internet-based service workload. We likewise depict a meta-arrangement foundation to build up worldwide Cloud trades and advertise, and show a contextual analysis of outfitting 'Storage Clouds' for superior substance conveyance. At last, author finish up with the requirement for joining of contending IT ideal models to convey our 21st century vision [7].

Leakage-Aware Multiprocessor Scheduling this paper, leakage-aware planning heuristics are introduced that decide the best exchange off between these three methods: DVS, processor shutdown, and finding the ideal number of processors. Exploratory results got utilizing a public benchmark set of assignment charts and genuine parallel applications demonstrate that our methodology lessens the aggregate vitality utilization by up to 46% for tight due dates and by up to 73% for free due dates thought about to a methodology that just utilizes DVS. Author likewise think about the vitality devoured by our booking calculations to two supreme lower limits, one for the situation where all processors ceaselessly keep running at the same recurrence, and one for the situation where the processors can keep running at diverse frequencies and these frequencies might change after some time.

The outcomes demonstrate that the vitality decrease accomplished by our best approach is near these hypothetical limits [8].

Profit-drive schedule for cloud services with data access awareness this paper, authors address the compromise of these scheduling so as to clash targets service demands with the element production of service examples. In particular, author booking calculations endeavor to expand benefit inside the agreeable level of service quality indicated by the service buyer. Author's commitments incorporate (1) the improvement of an evaluating model utilizing processor-sharing for cloud,the use of this estimating model to composite services with reliance thought, (3) the advancement of two arrangements of service solicitation booking calculations, and (4) the advancement of a prioritization arrangement for data service planning to amplify the benefit of data service [9]

Energy and Performance Management of Green Data Centers this paper, author try to handle this deficiency by proposing a precise way to deal with amplify green server farm's benefit, i.e., income short cost. In such manner, authors unequivocally consider reasonable service level agreement (SLAs) that as of now exist between information focuses and their clients. This model additionally fuses different elements, for example, accessibility of neighborhood renewable force era at server farms and the stochastic way of server farms' workload. Moreover, authors propose a novel advancement based benefit expansion procedure for server farms for two diverse cases, without and with behind-the-meter renewable generators. Authors demonstrate that the figured advancement issues in both cases are arched projects; in this manner, they are tractable and fitting for down to earth execution. Utilizing different test information what's more, by means of PC reproductions, authors evaluate the execution of the proposed advancement based benefit expansion methodology and demonstrate that it fundamentally outflanks two practically identical vitality and execution administration calculations that are as of late proposed in the writing [10].

III. RELATED WORK

Bellare et al. [3] showed Data confidentiality by transforming the predictable message into unpredictable message. Introduces a key server as third party to generate the file tag for duplicate check. Bugiel et al. [6] provided an architecture consisting of twin clouds for secure outsourcing of data and arbitrary computations to an entrusted commodity cloud. CSU et al. [9] also addressed the problem and showed a secure convergent encryption for efficient encryption, without considering issues of the key-management and block-level Deduplication.

Anderson,Le Zhang. [1] Proposed backup solutions for fast and secure backups used an encrypted Deduplication algorithm. This algorithm supports client-end per-user encryption which is necessary for confidential personal data. It also supports a unique feature which allows immediate detection of common sub trees, avoiding the need to query the backup system for every file.

IV. EXISTING SYSTEM

In the existing 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 a file, the user needs to take this file and his own privileges as inputs. The user is able to find a duplicate for this if and only if there is a copy of this file and a matched privilege stored in cloud.

Symmetric encryption technique.

Symmetric encryption uses a common secret key to encrypt and decrypt information. The user needs to know private key. Less protect security. These

Deduplication systems cannot support differential authorization duplication check.

V. PROPOSED SYSTEM

In the proposing system, we .eliminating duplicate copies of repeating data and has been widely used in cloud storage to reduce the amount of storage space and save bandwidth. To protect the privacy of sensitive data while supporting Deduplication, the convergent encryption technique has been proposed to encrypt the data before outsourcing .To better protect data security, this paper makes the first attempt to formally address the problem of authorized data Deduplication.

CONVERGENT ENCRYPTION TECHNIQUE

A user derives a convergent key from each original data copy and encrypts the data copy with the convergent key. The key generation algorithm that maps a data copy to a convergent key. The symmetric encryption algorithm that takes both the convergent key and the data copy as inputs and then outputs a cipher text. The decryption algorithm that takes both the cipher text and the convergent key as inputs and then outputs the original data copy and the tag generation algorithm that maps the original data copy and outputs a tag.

Whenever someone wants to give information or take information from cloud they have to take permission i.e. authentication is done. If not a member they have to register first. Then the user will request private cloud to get a file token .Private cloud will issue file token and Convergent key generation takes place. With that key user will upload a file to the public cloud, then there will be a Deduplication system to check whether the file already exist or not. If the file already exists then the file will not upload in public cloud.

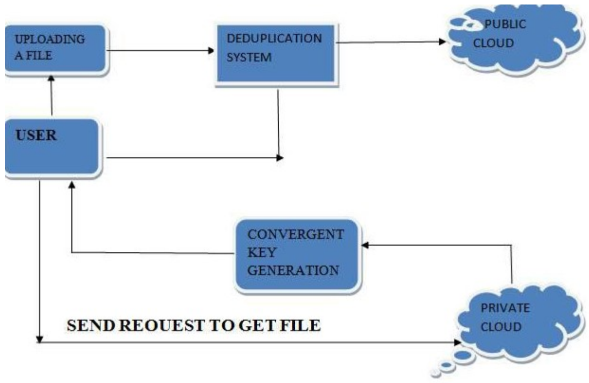
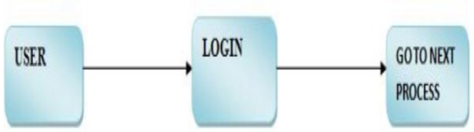


Figure 1. System architecture

AUTHENTICATION:

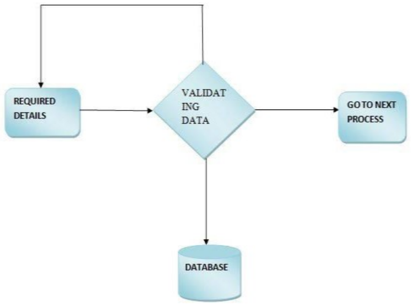
The process of identifying an individual usually based on a username and password. In security systems, Authentication merely ensures that the individual is who he or she claims to be, but says nothing about the access rights of the individual. In authentication module is used to security purpose. Here this module only for user, after registration user enter the username and password. This input is check into the database, whether input is correct or not. If input is correct then allow to next process otherwise consider as a non authenticated user.



a)Authentication

REGISTER:

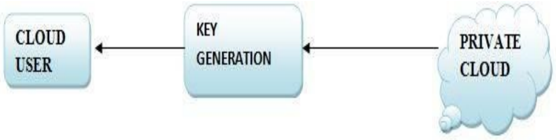
In this Module If he is a new user he needs to enter the required data to register the form and the data will be stored in server for future authentication purpose.



b) Register

CONVERGENT KEY GENERATION

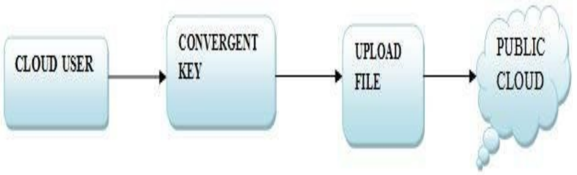
In this module, if user wants to upload a file user needs to get key from private cloud.



C) Convergent Key Generation

FILE UPLOADING

User can upload a file into the private cloud by using convergent key.



d) File Uploading

AUTHORIZED DUPLICATE CHECK SCHEME (ADS)

The public cloud performs duplicate check directly and tells the user if there is any duplicate. Public Cloud can store and retrieve file. De-duplication has a removing duplicate file. Its will find out duplicate file.

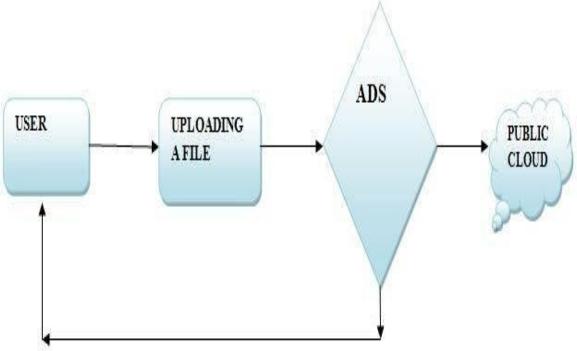


Figure 2. Authorized Duplicate Check Scheme

VI. APPLICATIONS

CtrlS Real Cloud:

The CtrlS Real Cloud has a multi-layered management model. The cloud controller server enables everything, from system architecture to VM root access, to be managed via the user interface and API. Real Cloud enables you to put up applications and manage them, all remotely and with utmost ease.

Cloud Layer Services:

Discover the promise of cloud, not the compromises. Cloud Layer includes virtual servers, remote storage and a robust content delivery network that leverage our core advantages and longtime leadership in automated, on-demand, self-managed infrastructure.

VII. CONCLUSION

In this paper, we have assessed distinctive information Deduplication procedures over scrambled information that is utilized as a part of the cloud processing for secure information storage. Customary encryption makes Deduplication unthinkable on account of the randomization property of encryption. As of late, a few Deduplication plans are proposed to understand this issue by enabling every proprietor to have a similar encryption key for similar information. Concurrent encryption has diverse encryption variations for secure Deduplication. However, CE experiences security blemishes with respect to label consistency and ownership renouncement. Besides, numerous plans couldn't accomplish secure access control under unique condition. Subsequently, very little work has yet been done to address dynamic ownership administration and its related security issue. Along these lines the proposed plot guarantees that exclusive approved access to the mutual information is conceivable, which is thought to be the most vital test for productive and secure cloud storage services in the earth where ownership changes progressively.

VIII. REFERENCES

1. 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.

2. M. Bellaire, S. Keelveedhi, and T. Ristenpart, “Dupless: Serveraided encryption for deduplicated storage,” in Proc. 22nd USENIX Conf. Sec. Symp., 2013, pp. 179–194.
3. 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.
4. 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.
5. 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.
6. 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.
7. 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.
8. D. Ferraiolo and R. Kuhn, “Role-based access controls, ” in Proc. 15th NIST-NCSC Nat. Comput. Security Conf., 1992, pp. 554–563.
9. J. Xu, E.-C. Chang, and J. Zhou. “Weak leakage-resilient client-side deduplication of encrypted data in cloud storage”. In ASIACCS, pages 195–206, 2013.