IPHV8I1074X
International Journal Of Advanced Research and Innovation -Vol.8, Issue .I
ISSN Online: 2319 – 9253
Print: 2319 – 9245

SECURE DEDUPLICATION WITH EFFICIENT AND RELIABLE CONVERGENT KEY MANAGEMENT IN CLOUD STORAGE

[#1]Y.SHAMANTH RAO, M.Tech Student,
[#2]N.THIRUPATHI, Associate Professor,
Department of CSE,
JYOTHISHMATHI INSTITUTE OF TECHNOLOGICAL SCIENCES,KARIMNAGAR,T.S,INDIA.

ABSTRACT: Secure deduplication is a technique for eliminating duplicate copies of storage data, and provides security to them. To reduce storage space and upload bandwidth in cloud storage deduplication has been a well-known technique. For that purpose convergent encryption has been extensively adopt for secure deduplication, critical issue of making convergent encryption practical is to efficiently and reliably manage a huge number of convergent keys. The basic idea in this paper is that we can eliminate duplicate copies of storage data and limit the damage of stolen data if we decrease the value of that stolen information to the attacker. This paper makes the first attempt to formally address the problem of achieving efficient and reliable key management in secure deduplication. We first introduce a baseline approach in which each user holds an independent master key for encrypting the convergent keys and outsourcing them. However, such a baseline key management scheme generates an enormous number of keys with the increasing number of users and requires users to dedicatedly protect the master keys. To this end, we propose Dekey, User Behavior Profiling and Decoys technology. Dekey new construction in which users do not need to manage any keys on their own but instead securely distribute the convergent key shares across multiple servers for insider attacker. As a proof of concept, we implement Dekey using the Ramp secret sharing scheme and demonstrate that Dekey incurs limited overhead in realistic environments. User profiling and decoys, then, serve two purposes. First one is validating whether data access is authorized when abnormal information access is detected, and second one is that confusing the attacker with bogus information. We posit that the combination of these security features will provide unprecedented levels of security for the deduplication in insider and outsider attacker.

Keywords: *Secure deduplication, Dekey, User Behavior Profiling, Decoy Technology.*

I.INTRODUCTION

Cloud computing is model of the distribution of the information services in which the resources are the retrieved from the web through some of the interfaces and applications, instead forming direct connections to the server. The fast expansion in information sources has mandatory for the users to make use of some of the storage systems for storing their secret data.

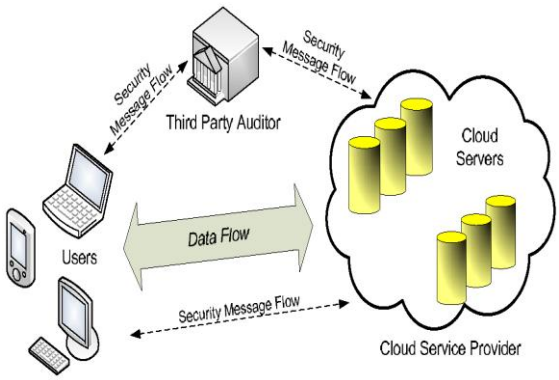

Fig. 1: The architecture of cloud data storage service

To enable privacy-preserving public auditing for cloud data storage under the aforementioned model, our protocol design should achieve the following security and performance guarantee:

1) Public auditability: to allow TPA to verify the correctness of the cloud data on demand without retrieving a copy of the whole data or introducing additional on-line burden to the cloud users.[2]

2) Storage correctness: to ensure that there exists no cheating cloud server that can pass the audit from TPA without indeed storing users' data intact.

3) Privacy-preserving: to ensure that there exists no way for TPA to derive users' data content from the information collected during the auditing process.

4) Batch auditing: to enable TPA with secure and efficient auditing capability to cope with multiple auditing delegations from possibly large number of different users simultaneously.[3]

5) Lightweight: to allow TPA to perform auditing with minimum communication and computation overhead.

Cloud storage systems provide the management of the ever increasing quantity of data by keeping in mind factors like reduce occupation storage space and the network bandwidth. To make the scalable and consistent management of the data in the cloud computing, deduplication technique plays an important role. Data

IPHV8I1074X
International Journal Of Advanced Research and Innovation -Vol.8, Issue .I
ISSN Online: 2319 – 9253
Print: 2319 – 9245

deduplication also helps to improve the results in efficiency term and searches are quicker. Data deduplication may happen as file level deduplication or as block level data deduplication. Instead of maintaining numerous duplicate copies of file or the data with alike content, deduplication senses and remove the redundant data by keeping original physical copy. Data deduplication is a technique of eliminate duplicate copies of data, and it is used in cloud storage to reduce storage space and bandwidth. An arising challenge is to perform secure deduplication in cloud storage even if convergent encryption is extensively adopted for secure deduplication; a critical issue is that making of convergent encryption practical to manage a huge number of convergent keys efficiently and reliably.

II. RELATED WORK

We show the way to style secure deduplication systems with higher reliableness in cloud computing. We introduce the distributed cloud storage servers into deduplication systems to produce higher fault tolerance. To more shield knowledge confidentiality, the key sharing technique is employed, that is additionally compatible with the distributed storage systems. in additional details, a file is first split and encoded into fragments by victimisation the technique of secret sharing, rather than coding mechanisms. These shares are going to be distributed across multiple independent storage servers. moreover, to support deduplication, a brief cryptologic hash price of the content also will be computed and sent to every storage server because the fingerprint of the fragment hold on at every server. solely the information owner UN agency 1st uploads the information is needed to reckon and distribute such secret shares, while all following users UN agency own an equivalent knowledge copy do not got to reckon and store these shares any longer. To recover knowledge copies, users should access a minimum number of storage servers through authentication and obtain the key shares to reconstruct the information. In other words, the key shares of knowledge can solely be accessible by the approved users UN agency own the corresponding knowledge copy.

Data Deduplication

Data deduplication is a technique for eliminating duplicate copies of data, and has been widely used in cloud storage to reduce storage space and upload bandwidth. Promising as it is, an arising challenge is to perform secure deduplication in cloud storage. Although convergent encryption has been extensively adopted for secure deduplication, a critical issue of making convergent encryption practical is to efficiently and reliably manage a huge number of convergent keys. One critical challenge of today‟s cloud storage services is the management of the ever-increasing volume of data. To make data management scalable deduplication we are use convergent Encryption for secure deduplication services.

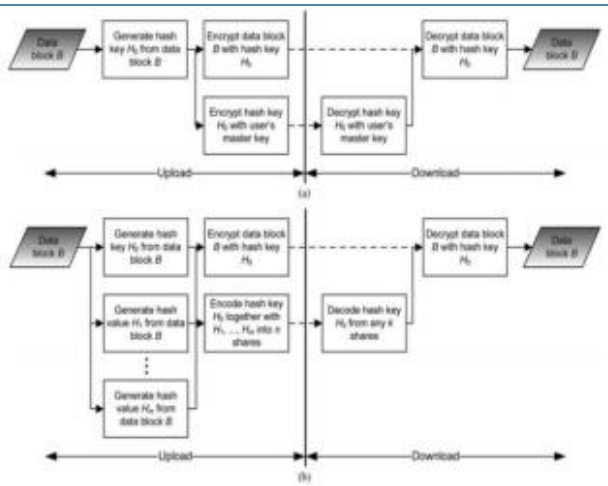

Fig 2: Secure deduplication
(a)Flow diagram keeping hash key
(b) Flow diagram of Dekey keeping hash key with RSSS.

User Behavior Profiling

By monitoring data access in the cloud and detect abnormal data access patterns User profiling is a wellknown Technique that can be applied here to model how, when, and how much a user accesses their information in the Cloud. Such „normal user‟ behavior can be continuously checked to determine whether abnormal access to a user‟s information is occurring. This method of behavior-based security is commonly used in fraud detection applications. Such profiles would naturally include volumetric information, how many documents are typically read and how often. We monitor for abnormal search behaviors that exhibit deviations from the user baseline the correlation of search behavior anomaly detection with trap-based decoy files should provide stronger evidence of malfeasance, and therefore improve a detector‟s accuracy.

Decoy Technology:

Decoy technology is the technology which is providing the decoy information to the unauthorized user or the attacker. Decoy technologies for example honey pot, or the generating the useless data files on the demand of the system to do attack against the attacker. Using this technique the original information gets changed in unexpected format so that the ex-filtering of the document or information is becomes impossible. This technology may be integrated with user behavior profiling technology to secure a user‟s information in the Cloud. Whenever abnormal access to a cloud service is noticed, decoy information may be returned by the Cloud and delivered in such a way as to appear completely legitimate and normal. The true user, who is the owner of the information, would readily identify when decoy information is being returned by the Cloud, and hence could alter the Cloud‟s responses through a variety of means, such as challenge questions, to inform the Cloud security system that it has inaccurately detected an unauthorized access. In the case where the access is correctly identified as an unauthorized access, the

Cloud security system would deliver unbounded amounts of bogus information to the adversary, thus securing the user‟s true data from unauthorized disclosure.

The traditional deduplication ways can not be directly extended and applied in distributed and multi-server systems. to clarify more, if identical short worth is stored at a unique cloud storage server to support a duplicate check by employing a ancient deduplication method, it cannot resist the collusion attack launched by multiple servers. In alternative words, any of the servers can acquire shares of the info keep at the opposite servers with identical short worth as proof of possession. moreover, the tag consistency, that was initial formalized by [5] to forestall the duplicate/ciphertext replacement attack, is taken into account in our protocol. in additional details, it prevents a user from uploading a maliciously-generated ciphertext such its tag is that the same with another honestly-generated ciphertext. to realize this, a settled secret sharing technique has been formalized and utilized. To our information, no existing work on secure deduplication will properly address the responsibleness and tag consistency downside in distributed storage systems. This paper makes the subsequent contributions. • Four new secure deduplication systems ar planned to provide economical deduplication with high reliability for file-level and block-level deduplication, respectively. the key rending technique, instead of ancient secret writing ways, is employed to protect knowledge confidentiality. Specifically, data are split into fragments by exploitation secure secret sharing schemes and keep at totally different servers. Our proposed constructions support each file-level and block-level deduplications .

• Security analysis demonstrates that the planned deduplication systems are secure in terms of the definitions specified in the planned security model. In more details, confidentiality, irresponsibleness and integrity can be achieved in our planned system. Two kinds of collusion attacks are thought-about in our solutions. These are the collusion attack on the info and also the collusion attack against servers. especially, the data remains secure notwithstanding the oppose controls a restricted range of storage servers.

• We tend to implement our deduplication systems exploitation the Ramp secret sharing theme that permits high irresponsibleness and confidentiality levels. Our analysis results demonstrate that the new planned constructions are economical and also the redundancies are optimized and comparable the opposite storage system supporting identical level of responsibleness. In previous deduplication systems cannot support differential authorization duplicate check, that is vital in several and applications. In such a licensed deduplication system, every user is issued a group of privileges throughout system data formatting.

III. PROPOSED SYSTEM

This section is devoted to the definitions of the how system model and security threats are work. In deduplication system two types of entities are their one is user and another is s cloud storage service provider(S-CSP).In this system model, to save the bandwidth for data uploading and storage space for data storing in the cloud both client and server side deduplication are supported. In order to save bandwidth of the uploading data and reliable management, the data will be moved to the s cloud server (S-CSP).This technique will be used for the storing only one copy of the same file in the cloud. The user is an entity that wants to store data on the outside outsource data storage and access the data later when user wants. In a cloud storage system deduplication, the user only uploads unique data or but does not upload any same copy of the file to save the upload bandwidth. Furthermore, the main thing is required by users to provide higher reliability in the system, As part of constructing our security model, it is important to establish a consistent notation. For achieving confidentiality and integrity to storing data in the cloud, the data deduplication system has been proposed. The main objective of this system is avoid duplicate storage of the data across distributed storage servers. To keep the confidentiality of the data and integrity of the data, our new constructions utilize the data splitting technique to divide the data into chunks. These chunks will then be distributed across multiple storage servers. In this paper we try to minimize the storage of the system. 3. Building Blocks A. S-CSP. The S-CSP is the storage cloud server provider service that provides the outsourcing data storage for the users. In the data deduplication system, when users wants to store the same data , the S-CSP will only store a single copy of these files and store only exclusive data. 3.1 The File-level Distributed Deduplication System. To support better duplicate check, tags for each chunk of the file which will be allocated and computed are sent to S-CSPs. To avoid collusion attack the S-CSPs, the tags stored at different distributed storage servers are logically independent and different. We now describe the details of the construction as follows.

A. File Upload. To upload a file F on the storage server, the user interacts with S-CSPs to perform the data deduplication. The user first calculates and sends the file tag ϕF = TagGen (F) to Storage-Cloud Server Provider for the file duplicate check. If a duplicate is found, the user processes and sends the result ϕF; idj = TagGen′ (F, idj) to the j-th server with identity idj through the secure channel for $1 \leq j \leq n$. Therefor logic behind is that an index j is to avoid the server from gaining the shares of other S-CSPs for the same data in a file or block, which will be expressed in

detail in the security analysis. If ϕF;idj same as the metadata stored with ϕF , the user will supplied a pointer for the chunk stored at server idj .Otherwise, if no duplication is found, the user will perform the computation as follows. He runs the secret sharing algorithm SS on F to get the {cj} = Share (F), where cj is the j-th chunk of F. He also processes ϕF;idj = TagGen′(F, idj), which provide the tag for the jth S-CSP. Finally, the user get uploads the set of values {ϕF , cj , ϕF;idj} to the S-CSP with identity idj through a secure channel. The S-CSP stores these data values and pointer get back to the user for its regular storage.

B. File Download. To download a file F, the user first get the secret shares {cj} of the data or file from k out of n distributed storage servers. Respectively, the user sends the pointer of F to k out of n Storage -Cloud Service Providers. After getting enough shares, the user rebuild file F by using this algorithm strategy of Recover ({cj}). This technique provides fault tolerance and grant the user to remain available even if any limited part of storage servers fail.

3.2 The Block-level Distributed Deduplication System In this section, we express that how to achieve the fine-grained block-level distributed deduplication. In a block-level deduplication system, the user also needs to firstly perform the file-level deduplication before uploading his file. If no repeated data is found, the user divides this file into blocks and performs block-level deduplication. The system setup and the file-level deduplication system both are same, except the block size parameter will be added additionally. Next, File Upload and File Download, this are the two method used in this algorithms. To upload a file F on distributed storage server, the user first performs the file-level deduplication by sending request? F to the storage servers.

Whenever, duplication is occur in a file, then user will perform file-level deduplication on that file F. Otherwise, user directly perform block-level deduplication on that file F as follows- Firstly File F is divide in into chunks {Ci} where i = 1, ... n). for each chunk Ci, computing? Ci = TagGen(Ci) for performing block level duplication, When the content of block level and file level are same then file is overlapped with the block Ci. Upon receiving block tags {? Ci}, the server with identity idj computes a block signal vector sCi for each i.

i) If sCi=1, the user further computes and sends? Ci;j = TagGen'(Ci, j) to the individuality of S-CSP with idj . If it also same as the corresponding tag stored, S-CSP returns a block pointer of Ci to the user. Then, the user keeps the block pointer of Ci and does not need to upload Ci.

ii) If sCi=0, the user runs the secret sharing algorithm SS over Ci and gets {bij} = Share (Ci), where bij is the j-th secret share of Ci. The user also computes ?Ci;j for 1 = j = n and uploads the set of values {?F , ?F;idj , bij , ?Ci;j} to the server idj via a secure channel. The corresponding pointers back to the user through S-CSP. File Download. To download a file F = {Ci}, the user first downloads the secret shares {bij} of all the blocks Ci in F from k out of n S-CSPs. Specifically, the user sends all the pointers for Ci to k out of n servers. After collecting all the shares, the user reconstructs all the fragments Ci using the algorithm of Recover ({·}) and gets the file F = {Ci}.In this paper, the data which is present in the file is uploaded by the user on the distributed server. After that server compare the chunks of the file by distributing them on to the servers. If any chunk of the file is matches with uploaded chunk of the file then, it will be directly discarded that particular chunk of the file. Using this technique, it will reduces the size of the servers storage and achieve the good reliability.

IV. SYSTEM METHODOLOGY

In our previous data deduplication systems, the non-public cloud is bothered as a proxy to allow knowledge owner/users to firmly perform duplicate talk over with differential privileges. Such style is sensible and has attracted lush attention from researchers. The data homeowners exclusively source their information storage by utilizing public cloud whereas the data operation is managed privately cloud. data deduplication is one among necessary data compression techniques for eliminating duplicate copies of repetition knowledge, and has been wide used in cloud storage to chop back the quantity of cabinet house and save system of measurement. To safeguard the confidentiality of sensitive data whereas supporting deduplication, Cloud computing provides ostensibly unlimited ‚virtualized' resources to users as services across the whole internet, whereas activity platform and implementation details. Today"s cloud service suppliers offer every extraordinarily offered storage and massively parallel computing resources at comparatively low costs. As cloud computing becomes rife, Associate in Nursing increasing amount of knowledge is being keep inside the cloud and shared by users with nominal privileges, that define the access rights of the keep data

Secure Data Deduplication

Data deduplication is a technique for eliminating duplicate copies of data, and has been widely used in cloud storage to reduce storage space and upload bandwidth. Promising as it is, an arising challenge is to perform secure deduplication in cloud storage. Although convergent encryption has been extensively adopted for secure deduplication, a critical issue of making convergent encryption practical is to efficiently and reliably manage a huge number of convergent keys. One critical challenge of today"s cloud storage services is the management of the ever-increasing volume of data. To make data management scalable deduplication we are use convergent Encryption for secure deduplication services.

IPHV8I1074X
International Journal Of Advanced Research and Innovation -Vol.8, Issue .I
ISSN Online: 2319 – 9253
Print: 2319 – 9245

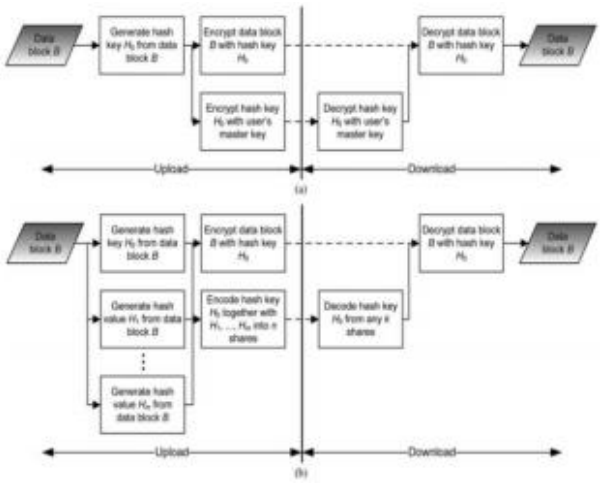

Fig 2: Secure deduplication
(a)Flow diagram keeping hash key
(b) Flow diagram of Dekey keeping hash key with RSSS.

User Behavior Profiling

By monitoring data access in the cloud and detect abnormal data access patterns User profiling is a wellknown Technique that can be applied here to model how, when, and how much a user accesses their information in the Cloud. Such „normal user" behavior can be continuously checked to determine whether abnormal access to a user"s information is occurring. This method of behavior-based security is commonly used in fraud detection applications. Such profiles would naturally include volumetric information, how many documents are typically read and how often. We monitor for abnormal search behaviors that exhibit deviations from the user baseline the correlation of search behavior anomaly detection with trap-based decoy files should provide stronger evidence of malfeasance, and therefore improve a detector"s accuracy.

Decoy Technology:

Decoy technology is the technology which is providing the decoy information to the unauthorized user or the attacker. Decoy technologies for example honey pot, or the generating the useless data files on the demand of the system to do attack against the attacker. Using this technique the original information gets changed in unexpected format so that the ex-filtering of the document or information is becomes impossible. This technology may be integrated with user behavior profiling technology to secure a user"s information in the Cloud. Whenever abnormal access to a cloud service is noticed, decoy information may be returned by the Cloud and delivered in such a way as to appear completely legitimate and normal. The true user, who is the owner of the information, would readily identify when decoy information is being returned by the Cloud, and hence could alter the Cloud"s responses through a variety of means, such as challenge questions, to inform the Cloud security system that it has inaccurately detected an unauthorized access. In the case where the access is correctly identified as an unauthorized access, the Cloud security system would deliver unbounded amounts of bogus information to the adversary, thus securing the user"s true data from unauthorized disclosure.

Block-Level Deduplication

In a block-level deduplication system, the user also needs to firstly perform the file-level deduplication before uploading his file. If no duplicate is found, the user divides this file into blocks and performs block-level deduplication. Block level deduplication, which discovers and removes redundancies between data blocks. The file can be divided into smaller fixed-size or variable-size blocks. Using fixed size blocks simplifies the computations of block boundaries, while using variable-size blocks (e.g., based on Rabin fingerprinting) provides better deduplication efficiency.

V. CONCLUSION

We projected the distributed deduplication systems to improve the reliableness of information whereas achieving the confidentiality of the users" outsourced knowledge while not Associate in nursing encryption mechanism. Four constructions were projected to support file-level and fine grained block-level data deduplication. the safety of tag consistency and integrity were achieved. We enforced our deduplication systems victimization the Ramp secret sharing theme and demonstrated that it incurs tiny encoding/decoding overhead compared to the network transmission oveoverhead in regular upload/download operations. We can achieve this with the help of preventive disinformation attack. We posit that secure deduplication services can be implement given additional security features insider attacker and outsider attacker by using the detection of masquerade activity.

REFERENCES

[1]M. Bellare, A. Desai, E. Jokipii, and P. Rogaway. A Concrete Security Treatment of Symmetric Encryption: Analysis of the DES Modes of Operation. Proceedings of the 38th Symposium on Foundations of Computer Science, IEEE, 1997.

[2] Ayushi "A Symmetric Key Cryptographic Algorithm " International Journal of Computer Applications (0975 - 8887) ©2010 Volume 1 – No. 15

[3] Abdul Wahid Soomro, Nizamuddin, Arif Iqbal Umar, Noorul Amin." Secured Symmetric Key Cryptographic Algorithm for Small Amount of Data" 3rd International Conference on Computer & Emerging Technologies (ICCET 2013).

[4] A. Rahumed, H. Chen, Y. Tang, P. Lee, and J. Lui. A secure cloud backup system with assured deletion andversion control. In Parallel Processing Workshops

(ICPPW), 2011 40th International Conference on, pages160-167 IEEE, 2011.

[5]Z. Wilcox-O'Hearn and B. Warner. Tahoe: The least-authority _lesystem. In Proceedings of the 4th ACM international workshop on Storage security and survivability, pages 21-26. ACM, 2008.

[6]S. P. Vadhan. On constructing locally computable extractors and cryptosystems in the bounded storage model. In D. Boneh, editor, CRYPTO 2003, volume 2729 of LNCS, pages 61-77. Springer, Aug. 2003.

[7].Jin Li, Yan Kit Li, Xiaofeng Chen, Patrick P. C. Lee, Wenjing Lou" A Hybrid Cloud Approach for Secure Authorized Deduplication" IEEE Transactions On Parallel And Distributed System VOL:PP NO:99 YEAR 2013.

[8]M. Ben-Salem and S. J. Stolfo, "Modeling user search-behavior for masquerade detection," in Proceedings of the 14th International Symposium on Recent Advances in Intrusion Detection . Heidelberg: Springer, September 2011, pp. 1–20.

[9] Salvatore J. Stolfo, Malek Ben Salem and Angelos D. Keromytis "Fog Computing: Mitigating Insider Data Theft Attacks in the Cloud" IEEE Symposium On Security And Privacy Workshop (SPW) YEAR 2012

[10] I.Sudha1, A.Kannaki2, S.Jeevidha3" Alleviating Internal Data Theft Attacks by Decoy Technology in Cloud", International Journal of Computer Science and Mobile Computing, Vol.3 Issue.3, March- 2014, pg. 217-222.

[11] B. M. Bowen and S. Hershkop, "Decoy Document Distributor: http://sneakers.cs.columbia.edu/ids/fog/," 2009. [Online]. Available: http://sneakers.cs.columbia.edu/ids/FOG/

[12] Jin Li, Xiaofeng Chen, Mingqiang Li, Jingwei Li, Patrick P.C. Lee, and Wenjing Lou "Secure Deduplication with Efficient and Reliable Convergent Key Management" IEEE Transactions On Parallel And Distributed Systems, VOL. 25, NO. 6, JUNE 2014..

[13] Mr N.O.Agrawal, Prof. S.S.Kulkarni"Secure Deduplication and Data Security with efficient and reliable CEKM" IJAIEM Transition On paral