iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 3 **Issue:** II **Month of publication:** February 2015

DOI:

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Detailed Investigation on a Hybrid Cloud Approach for Secure Authorized Deduplication

A.Abdul Samadhu[#1], J.Rambabu[*2], R.Pradeep Kumar[#3], R.Santhya[#4]

[#1234]*kalaignar karunanidhi Institute of Technology*

Abstract*—Data deduplication is one of important data compression techniques which is for eliminating duplicate copies of repeating data, and has been widely used in cloud storage in order to reduce the amount of storage space and save bandwidth. To protect the confidentiality of sensitive data while supporting deduplication, the convergent encryption technique has been proposed to encrypt the data before outsourcing. To better protect data security, paper's makes the first attempt to formally address the problem of authorized data deduplication. Different from traditional deduplication systems, the differential privileges of users are further considered in duplicate check besides the data itself. Also present several new deduplication constructions supporting authorized duplicate check in hybrid cloud architecture. Security analysis demonstrates that the proposed scheme is secure in terms of the definitions specified in the proposed security model.

Keywords*— Deduplication, authorized duplicate check, confidentiality, hybrid cloud, secure

I. INTRODUCTION

Cloud computing provides seemingly unlimited "virtualized" resources to users as services across the whole Internet, while hiding platform and implementation details. Today's cloud service providers offer both highly available storage and massively parallel computing resources at relatively low costs. As cloud computing becomes prevalent, an increasing amount of data is being stored in the cloud and shared by users with specified privileges, which define the access rights of the stored data. One critical challenge of cloud storage services is the management of the ever-increasing volume of data. To make data management scalable in cloud computing, deduplication has been a well-known technique and has attracted more and more attention recently. 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 can also be applied to network data transfers to reduce the number of bytes that must be sent. Instead of keeping multiple data copies with the same content, deduplication eliminates redundant data by keeping only one physical copy and referring other redundant data to that copy. 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 inside and outside attacks. Traditional encryption, while providing data confidentiality, is incompatible with data deduplication. Specifically, traditional encryption requires different users to encrypt their data with their own keys. Thus, identical data copies of different users will lead to different ciphertext, making deduplication impossible. Convergent encryption [8] has been proposed to enforce data confidentiality while making deduplication feasible. It encrypts/decrypts a data copy with a convergent key, which is obtained by computing the cryptographic hash value of the content of the data copy. After key generation and data encryption, users retain the keys and send the ciphertext to the cloud. Since the encryption operation is deterministic and is derived from the data content, identical data copies will generate the same convergent key and hence the same ciphertext. 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. Thus, convergent encryption allows the cloud to perform deduplication on the ciphertext and the proof of ownership prevents the unauthorized user to access the file.

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. For example, in a company, many different privileges will be assigned to employees. In order to save cost and efficiently management, the data will be moved to the storage server provider (SCSP) in the public cloud with specified privileges and the deduplication technique will be applied to store only one copy of the same file. Because of privacy consideration, some files will be encrypted and allowed the duplicate check by employees with specified privileges to realize the access control. 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. It seems to be contradicted if we want to realize both deduplication and differential authorization duplicate check at the same time.

II. CONVERGENT ENCRYPTION

Convergent encryption provides data confidentiality in deduplication. A user (or data owner) derives a convergent key from each original data copy and encrypts the data copy with the convergent key. In addition, the user also derives a *tag* for the data copy, such that the tag will be used to detect duplicates. Here, we assume that the tag correctness property holds, i.e., if two data copies are the same, then their tags are the same. To detect duplicates, the user first sends the tag to the server side to check if the identical copy has been already stored. Note that both the convergent key and the tag are independently derived, and the tag cannot be used to deduce the convergent key and compromise data confidentiality. Both the encrypted data copy and its corresponding tag will be stored on the server side. Formally, a convergent encryption scheme can be defined with four primitive functions:

KeyGenCE(M) *!* K is the key generation algorithm that maps a data copy M to a convergent key K;
EncCE(K,M) *!* C is the symmetric encryption algorithm that takes both the convergent key K and the data copy M as inputs and then outputs a ciphertext C;
DecCE(K,C) *!* M is the decryption algorithm that takes both the ciphertext C and the convergent key K as inputs and then outputs the original data copy M; and
TagGen(M) *!* $T(M)$ is the tag generation algorithmthat maps the original data copy M and outputs a tag $T(M)$.

III.SECURITY PROOFS FOR IDENTITY BASED IDENTIFICATION AND SIGNATURE SCHEMES

In this paper, the author provided both the security proofs or attacks for a large number of identity-based identification and signature schemes defined either explicitly or implicitly, underlying these is a framework that on one hand helps to explore these schemes and how it is derived and on the other hand it enables. In this paper, the author discussed about IBI (identity based identification) scheme and IBS (identity based signature).In IBI scheme the author said that there is an authority containing public key and a master secret key. This authority can given to a user with a secret key based on the identity . In case of IBS scheme, it is similar expect that the user signs message, than identifying itself and checking of the signature needs knowledge only of the identity of the signer and the master public key.

IV.REVDEDUP: REVERSE DEDUPLICATION STORAGE SYSTEM OPTIMIZED FOR READS TO LATEST BACKUPS

In this paper the author said that increasing the backup storage for an increasing volume of virtual machine (VM) images is an important issue in virtualization environment though the deduplication eliminates the duplicates for VM image storage it in turn will introduces fragmentation which will degrade the read performance. Thus to overcome this, the author proposed RevDedup which is a deduplication system that optimizes reads to latest VM. Image backups using a novel idea called reverse deduplication. This RevDedup removes duplicates from old data that shifts fragmentation to old data by keeping the layout of new data.

V. PRIVATE DATA DEDUPLICATION PROTOCOLS IN CLOUD STORAGE

In this paper, the author deals with a new concept that is said to be private data deduplication protocols which is been introduced and formalized in the content of two –party computations in this paper, the author shown that the proposed private data deduplication protocol will be secure in the simulation based framework by assuming that the given hast function is collision-resilient.

VI. BOOSTING EFFICIENCY AND SECURITY IN PROOF OF OWNERSHIP FOR DE-DUPLICATION

In this paper, the author describes the deduplication techniques. It is used for reducing the storage needed by service provider and it is based on the intuition. The same content will be stored for several reasons. Hence only single copy is enough to store. Deduplication states that to avoiding having to sore the data several times. The large sets of data are often exhibit high reductancy. In this paper, novel security protocols are present to implement the proof of ownership.

VII. A NEW APPROACH TO ARCHIVAL STORAGE

In this paper the author describes the network storage called Venti .Each and every hash of block's contents act as block identifier. In this the duplicate copy and consumption of storage is reduced. Constructing the varieties of storage application building block is called Venti and this is also describes the design and implementation of an archival storage. The main goal of Venti is to provide the write –once archival repository and this is shared by the multi client machines and applications. The Venti is a block-level network storage system intended for archival data. For storing a collected futes and directories as a single object the vac application is used with vac the contents of the selected data's are stored as a tree of blocks on a venti server. The Vic writes each files as a separate collection of venti block .so the writes –once model and duplicate copies of a block will makes venti a useful storage application.

VIII. A SECURE CLOUD BACKUP SYSTEM WITH ASSURED DELETION AND VERSION CONTROL

In this paper, the author describes the cloud storage and secure backup system so the author presents the "Fade Version" for secure cloud backup system that provide the security layer on top of today's cloud storage services .The fade version follows the std version controlled backup design which eliminates the storage of redundant data among the various versions of backup in this paper the process of concept Amazon S3 was implemented thus shows the fade version. Adds lower performance overhead over a traditional cloud backup services that does not support assured deletion. The cloud computing system is one of the emerging service model and it will provides computation and storage classes resources on the network .the enterprises can often required to remotely achieve this data in the case of system software or hardware and delivers .Assured deletion to provide cloud clients reliably destroying their data backups upon the requests. The fade version providing assured version control of data backup's .the performance of fade version is the additional storage of cryptographic keys in data backups .the future work of this paper is to reducing the number of keys to be stored and managed.

IX. ROLE BASED ACCESS CONTROL MODELS

In this paper, the author describes the RBAC (Role Based Access Control). The permissions are associated to the roles and members of roles. The roles are related to the user groups in access control. It is originated with the multi user computer system. This greatly simplifies the management of permission. The roles are used in the different job functions and it is assigned by the user based on their qualification. The role is stable it will change usually less frequently. The role is created for to do perform specific task. The purpose of RBAC security administration and review. This administrative use of roles found in the modern networks OS. The research problem in this area is developing systematic approach to the design. So the future work is to develop the systematic methodology and analysis of constraints etc., many of these open issue will require an integrated an integrated approach for their resolution.

X. CONCLUSIONS

In this paper, the investigation is based on the notion of authorized data deduplication was proposed to protect the data security by including differential privileges of users in the duplicate check. We also discussed several new deduplication constructions that 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. Security analysis demonstrates that our schemes are secure in terms of

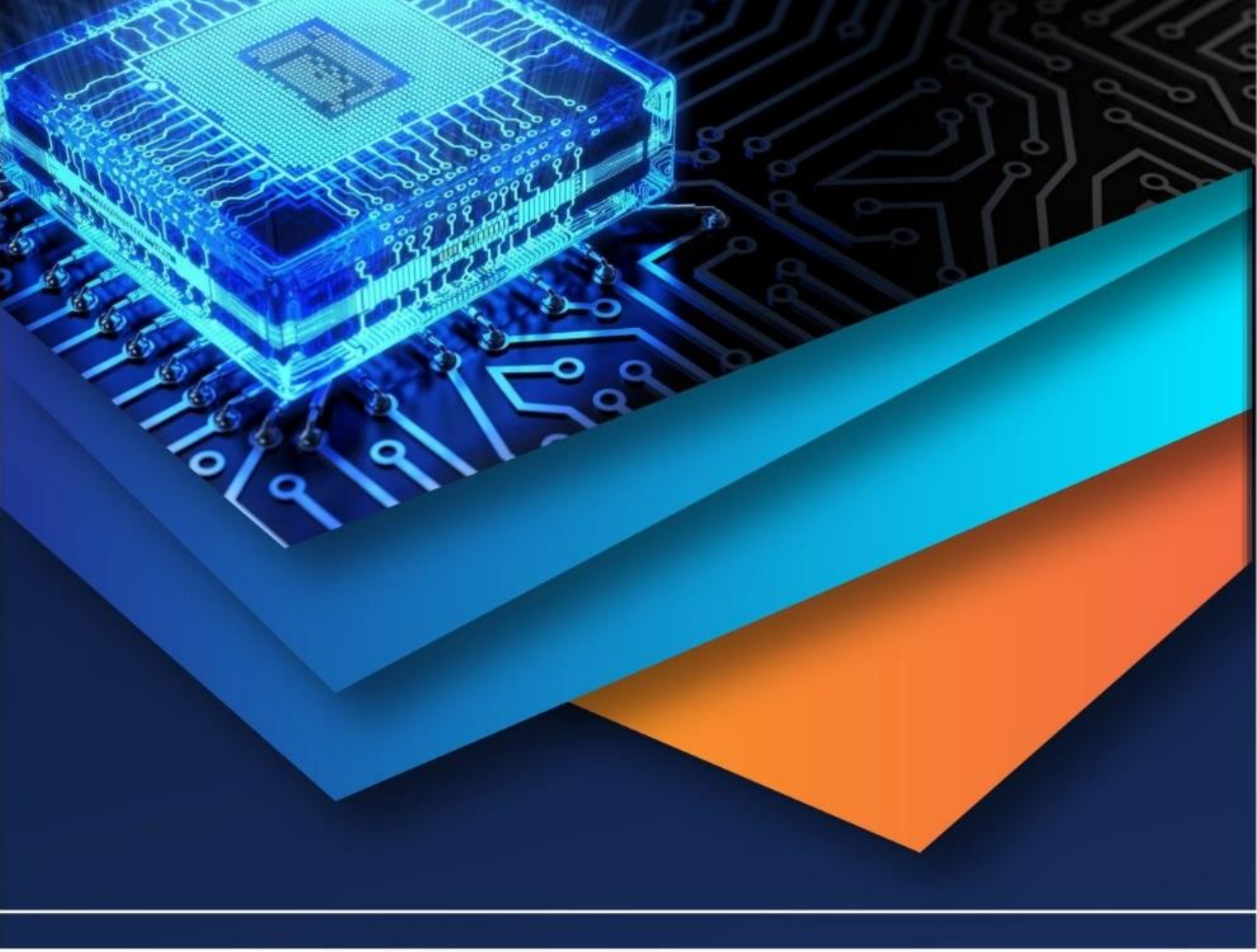
insider and outsider attacks specified in the proposed security model. We also planned to improve the dedepulication strategy by assigning the integrity check to highly authorized individuals.

XI.ACKNOWLEDGMENT

Reaching destination is not possible without a walk towards it. Of course yes, I would like to thank few people at this moment, who had been a great support for us to achieve it. With a start we would like to thank god and also we wish to thank Pongalur N Palanichamy, Founder Chairman, Mrs. Indhu Murugesan, Managing Trustee, Mr. N Anbalagan, Director, Mr. Mohandas Gandhi, Principal for their immense supervision providing all helpful needs for bringing this journal paper a successive completion. We would also like to thank Asst. Prof. Mr. Ganesh Moorthi, Kalaignar Karunanidhi Institute of Technology for his kind guidance and support for making a successful endeavour. Our sincere thanks for our colleagues in developing this project paper as a marvellous tribute.

REFERENCES

[1] OpenSSL Project. http://www.openssl.org/.
[2] P. Anderson and L. Zhang. Fast and secure laptop backups with encrypted de-duplication. In *Proc. of USENIX LISA*, 2010.
[3] M. Bellare, S. Keelveedhi, and T. Ristenpart. Dupless: Serveraided encryption for deduplicated storage. In *USENIX Security Symposium*, 2013.
[4] M. Bellare, S. Keelveedhi, and T. Ristenpart. Message-locked encryption and secure deduplication. In *EUROCRYPT*, pages 296– 312, 2013.
[5] M. Bellare, C. Namprempre, and G. Neven. Security proofs for identity-based identification and signature schemes. *J. Cryptology*, 22(1):1–61, 2009.
[6] M. Bellare and A. Palacio. Gq and schnorr identification schemes: Proofs of security against impersonation under active and concurrent attacks. In *CRYPTO*, pages 162–177, 2002.
[7] S. Bugiel, S. Nurnberger, A. Sadeghi, and T. Schneider. Twin clouds: An architecture for secure cloud computing. In *Workshop on Cryptography and Security in Clouds (WCSC 2011)*, 2011.
[8] 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.
[9] D. Ferraiolo and R. Kuhn. Role-based access controls. In *15th NIST-NCSC National Computer Security Conf.*, 1992.
[10] GNU Libmicrohttpd. http://www.gnu.org/software/libmicrohttpd/.
[11] 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.
[12] J. Li, X. Chen, M. Li, J. Li, P. Lee, andW. Lou. Secure deduplication with efficient and reliable convergent key management. In *IEEETransactions on Parallel and Distributed Systems*, 2013.
[13] libcurl. http://curl.haxx.se/libcurl/.
[14] C. Ng and P. Lee. Revdedup: A reverse deduplication storage system optimized for reads to latest backups. In *Proc. of APSYS*, Apr 2013.
[15] W. K. Ng, Y. Wen, and H. Zhu. Private data deduplication 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.
[16] R. D. Pietro and A. Sorniotti. Boosting efficiency and security in proof of ownership for deduplication. In H. Y. Youm and Y. Won, editors, *ACM Symposium on Information, Computer and Communications Security*, pages 81–82. ACM, 2012.
[17] S. Quinlan and S. Dorward. Venti: a new approach to archival storage. In *Proc. USENIX FAST*, Jan 2002.
[18] 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.
[19] R. S. Sandhu, E. J. Coyne, H. L. Feinstein, and C. E. Youman. Role-based access control models. *IEEE Computer*, 29:38–47, Feb 1996.
[20] J. Stanek, A. Sorniotti, E. Androulaki, and L. Kencl. A secure data deduplication scheme for cloud storage. In *Technical Report*, 2013.
[21] M. W. Storer, K. Greenan, D. D. E. Long, and E. L. Miller. Secure data deduplication. In *Proc. of StorageSS*, 2008.
[22] Z. Wilcox-O'Hearn and B. Warner. Tahoe: the least-authority filesystem. In *Proc. of ACM StorageSS*, 2008.
[23] 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.
[24] J. Yuan and S. Yu. Secure and constant cost public cloud storage auditing with deduplication. *IACR Cryptology ePrint Archive*, 2013:149, 2013.
[25] 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.
[26] Li, Jin, Y. Li, Xiaofeng Chen, Patrick PC Lee, and Wenjing Lou. "A Hybrid Cloud Approach for Secure Authorized Deduplication." (2014): 1-1.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY