ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 3, March 2015

A Survey on Secure Authorized Deduplication for Outsourced Data in Hybrid Cloud

Tamizh M[1], Elavarasi K[2]

Information Technology, IFET College of Engineering, Villupuram, India[1]

Assistant professor, Information Technology, IFET college of Engineering, Villupuram, India[2]

Abstract: Cloud Service Providers offer highly available storage and colossally computing resources at comparatively squat fee. The storage space provided may contain more number of repeated data. Inorder to reduce the storage space and eliminate the duplicates of data, Data Deduplication technique is being used. It is one of the method that eradicates duplicate copies of repeating data to reduce the storage space and save bandwidth. Inorder to protect the confidentiality of the data, Convergent Encryption Technique has been used. It outsource the data only after it is being encrypted using the convergent key. To protect the Data Security, Differential Authorized duplicate check is being used. The deduplication is done only by the authorized user. In traditional deduplication schemes, the disparity rights of users are further considered in duplicate check besides the data itself. The new deduplication constructions supports authorized duplicate check in a hybrid cloud architecture. The proposed system includes the VAST storage system that is used for storing the valuable and secret data. It thwarts the unauthorized user to steal the information from small pieces of the data. The hacker cannot attain the entire information from small part of information, since the data is stored in the form of manageable pieces of data.

Keywords: Convergent encryption technique, Authorized duplicate check, Private Cloud, Public Cloud, Deduplication, Privileges.

I.INTRODUCTION

In cloud computing, Data Compression technique is used for eliminating the duplicates of data.[13]A Hybrid cloud is the combination of private and public cloud. Private cloud is used for generating the private key and it is being handled by the user .Public cloud is used for storing the data that is outsourced.Currently,many of the cloud service providers offer extremely available repository and this consists of more number of repeated data[14].It is waste of memory, so the deduplication is used eradicate such repeated data by storing only one copy of the data. Cloud Computing is becoming ubiquitous, an aggregate volume of data is being deposited in the cloud and deposited data is pooled by users with quantified privileges, that outlines the admittance privileges to a precise users.[8]The privileges diverge for diverse applications such as role based privileges and time based privileges. The tricky of cloud storage services is the supervision of aggregate bulk of data on the cloud.

Data deduplication system eradicate the redundant data copies of repeating data in the cloud storage[7]. The system recovers the storage consumption and it can be applicable to network data transfer to reduce the number of bytes that must be sent. Data deduplication stretches lot of reimbursements, refuge and confidentiality anxieties ascend as the users' subtle data is liable to both inside and outside spasms. Profitable cloud storage services such as Drop box, Mozy, and Memopal, have been applying deduplication for user data to bar preservation cost .Data outsourcing raises security and privacy concerns.[11] Deduplication improves storage and bandwidth competence and is attuned with Convergent key management. Traditional encryption requires dissimilar users to encrypt their data with their own keys.[15] Edifice a reliable cloud computing setting is not enough, because fates endure to transpire and If evidence acquires gone. One needs to prepare for such fates. The basic idea is to limit the mutilation of embezzled data decrease the importance of that embezzled information to the invader.

II. LITERATURE SURVEY

The section focuses on the various process that is being done to reduce the storage space.[1] J. Stanek, A. Sorniotti, E.Androulaki,CloudDedup perform deduplication securely by encrypting the file.It assures confidentiality of data and block level deduplication. It remains secure to the component that implements block level deduplication. If the data is not encrypted, confidentiality cannot be attained. The convergent encryption technique is used to encrypt the data based on the convergent key. The convergent key derived is usually the hash value. Since the confidentiality and deduplication can be guaranteed by convergent encryption but it suffers from well-known weakness of Dictionary attacks. The Secret value S is determined inorder to overcome the well-known weakness. It is attuned with typical storing API's and translucent for cloud service providers. These process are based on securely performing deduplication and it is transparent to users.[2] W. K. Ng, Y. Wen, and H. Zhu,describes that deduplication is performed for the private cloud. It allows the user who holds the private data. The security of private cloud is based on the simulation-based framework. The private cloud is secure in the underlying hash function that is collision resilient and discrete logarithm. The collision-resilient muddle function is a polynomial time reckonable function H representing dualistic cords of whimsical length into

DOI 10.17148/IJARCCE.2015.4397

judiciously diminutive ones, so that it is computationally infeasible to determine any impact, that is any two different twines x and y for which H(x)=H(y).The discrete logarithm is rigid and the expurgation coding algorithm E can expurgation up to α-fraction of the jiffs in the manifestation of malevolent antagonists. These process is based on the private cloud deduplication and it does not provide high security.[3] C.Ng and P.Lee,RevDedup,a deduplication scheme enhances declaims to newest VM image snarls consuming an inkling baptized reverse deduplication. It eradicates duplicates from hoary data, thereby fluctuating breakup to old data although observance the draught of innovative data as chronological as imaginable. Deduplicationis used for backup storage.. The deduplication for backup storage focus on enhancing fingerprint indexing to triumph high snarl concert.RevDedup smears coarse-grained comprehensive deduplication to remunerate disk pursues over large size data units. It maintain high deduplication efficiency. It applies fine grained reverse deduplication, in which the data assignment as consecutive as conceivable for the latest version, while eradicating any redundant data of the old forms and stating it to the duplicate data of the latest form. It triumphs high deduplication efficiency, and in the meantime diminishes crumbling and triumphs high read enactment. The RevDedup eradicates the data from the old to new data. It does not provide that much security.[4]S. Bugiel,S.Nurnberger,A.Sadeghi, and T. Schneider ,determines the confident outsourcing of data and haphazard reckonings to an untrusted commodity cloud. The user converses with a trustworthy cloud (either a private cloud or numerous secure hardware sections)that encrypts and verifies the data stowed and maneuvers accomplished in the untrusted commodity cloud. The twin cloud is the combined form of trusted cloud and Commodity cloud. The security precarious maneuvers are accomplished by the Trusted Cloud. The enactment precarious maneuvers are accomplished on scrambled data by the Commodity Cloud. A client sorts consumption of the amenities offered by a cloud service provider to outsource its data and reckonings into the Commodity Cloud securely. The outsourced data is based on confidentiality and integrity protected. The exactness of the outsourced reckonings is to be verifiable by the client. It transfer the data based on the query.[5] S. Quinlan and S. Dorward, describes the archival data storage. It imposes a write-once policy, thwarting fortuitous or malevolent obliteration of data. It is a edifice wedge for fabricating a assortment of stowing solicitations such as coherent backup, corporeal backup, and Polaroid dossier. Venti ascertains data lumps by a hotchpotch of their innards. The collision-resistant muddle utility with adequately enormous output, since it determines the hash of a data block as irreplaceable. The inimitable hash is baptized the whorl of a block. The address for read and write operations.Vac of archives and almanacs as a solitary object, the functionality utilities tar and zip. The innards of nominated archives are stockpiled as tree of lumps on a Venti server. Thus this process provides security as well as the archive of data storage.

III. CONCLUSION

Data Deduplication eradicates the redundant data by storing only the single copies of data. It uses the convergent encryption technique to encrypt the data with the convergent key. It also provides Differential Authorized duplicate check, so that only authorized user with specified privileges can perform the duplicate check. The concept deduplications save the bandwidth and reduce the storage space. It also eradicates the duplicates of data in the cloud storage .The hybrid cloud provides lots benefits based on the confidentiality, authorized duplicate check. The VAST technique does not allow the unauthorized user to steal the entire information based on the single part of information they attained. It stores the data in the manageable pieces. Thus the cloud storage space as well as the healing an information is prevented.

IV. FUTURE WORK

In the future work,the VAST technique will first,prevent the attacker from stealing the valuable information.Secondly,it recover the corrupted data.Finally,it will recover the Byzantine attack file.

REFERENCES

[1] C. Ng and P. Lee. Revdedup: A reverse deduplication storage system optimized for reads to latest backups. In Proc. of APSYS, Apr 2013.

[2] J. Stanek, A. Sorniotti, E. Androulaki, and L. Kencl. A secure data deduplication scheme for cloud storage. In Technical Report, 2013

[3] 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.

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

[5] 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.

[6] 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.

[7] M. Bellare, S. Keelveedhi, and T. Ristenpart. Dupless: Server-aided encryption for deduplicated storage. In USENIX Security Symposium, 2013.

[8] M. Bellare, S. Keelveedhi, and T. Ristenpart. Message-locked encryption and secure deduplication. InEUROCRYPT,pages296– 312, 2013.

[9] M. Bellare, C. Namprempre, and G. Neven. Security proofs for identity-based identification and signature schemes. J. Cryptology, 22(1):1–61, 2009.

[10] 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.

[11] P. Anderson and L. Zhang. Fast and secure laptop backups with encrypted de-duplication. In *Proc. of USENIX LISA*, 2010.

[12] M. Bellare, S. Keelveedhi, and T. Ristenpart. Dupless: Server aided encryption for deduplicated storage. In *USENIX SecuritySymposium*, 2013.

[13] M.Bellare,S.Keelveedhi, and T.Ristenpart. Message-locked encryption and secure deduplication. In *EUROCRYPT*, pages 296–312, 2013.

[14] M. Bellare, C. Namprempre, and G. Neven. Security proofs for identity-based identification and signature schemes. *J. Cryptology*,22(1):1–61, 2009.

[15] 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.

DOI 10.17148/IJARCCE.2015.4397