ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 9, September 2015

A Review on an Approach to Minimize Storage Requirement in Cloud Using De-Duplication Mechanism

Naziya Tabassum, Prof. Roshani B. Talmale

M. Tech Student, Dept. of CSE, Tulsiramji Gaikwad-Patil College of Engineering and Technology Nagpur, India

Professor, Dept. of CSE, Tulsiramji Gaikwad-Patil College of Engineering and Technology Nagpur, India

ABSTRACT: Data de-duplication is one of the data compression techniques that can be used to eliminate the repetitive data and can be used in cloud computing architecture. For privacy and security point of view convergent technique has been used to encrypt data before outsourcing. But previous systems have the limitation of convergent encryption and in proposed system applied the techniques of cryptographic tuning to make the encryption more secure and flexible. Data de-duplication prohibits the storage of repetitive blocks and implements the pointer concept which basically puts the pointer to the existing blocks. Access control is provided into the application which allows the data owner the freedom of selecting users to have access to the published file. The integrity of data outsourced to the cloud is managed by the hash calculation of any content following the proof-of-ownership module. Proposed system calculate hash of the content on source and destination side and request the hash for the cloud side to predict the tampering of data. The expected analysis shows the improvement in execution time and development cost.

KEYWORDS: De-duplication, authorized duplicate check, confidentiality, public cloud.

I. INTRODUCTION

Cloud computing provide several attractive benefits for business, enterprises and end users. The benefits of cloud computing includes self-services provision that means the end user can use the resources for almost any type of workload on-demand, Elasticity that means company can scale as computing needs increase, Pay per use that means computing resources are measured as they are used or demand by the company. Cloud computing services can be public, private or hybrid.

Public cloud model is a third party provider that delivers the cloud services over the internet. Private cloud services are delivered from a business data centre to internet users. Hybrid cloud is a combination of public cloud services and private cloud services.

Cloud computing has been the most widely spread and recent technology that is being used in corporate model. As its flexibility of unlimited data storage and its access worldwide, it comes with a price. The cloud computing runs over “Pay As You Go” model and bills the storage cost to the company. Cloud services provider provides both highly availability and parallel computing resources at low cost. These resources can be dynamically configured to allow optimum resources utilization. As cloud computing grows the amount of data increases day by day and shred by user with specified privileges.

So it is most importance to first minimize the storage cost and eliminate the repetitive data. During the cost minimization, you have to also care about privacy about data and you cannot ignore the fact. As the amount of data increased the management of the increased data is difficult. To make data manageable in cloud computing a well-known technique de-duplication [1] and has attracted more attention recently. De-duplication basically used to improve the storage utilization and also applied in to network data transfer to improve the speed of data transmission. De-duplication eliminates the redundant data by keeping only one copy of data instead of keeping multiple copies of same data and referring other redundant data to that particular copy. Data de-duplication comes under two categories either file level or the block level. Data de-duplication brings a lot of benefits on security and privacy level because user sensitive data are susceptible to both inside and outside attacks.

The main objective of the proposed system is to provide a secure cloud computing architecture with storage as a service model. The proposed system understand the problem of outsourcing data to the cloud, is the privacy and security. To make data more secure proposed system need to provide encryption to the data to be outsourced. At the same time this need to keep the storage to be minimum and avoid duplication and redundancy. At the privacy level proposed system need to manage the access control mechanism to prohibit the unauthorized users to get access to the data.

II. RELATED WORK

In cloud computing security of data de-duplication has attracted much attention from research. A. Rahumed in [12] presents a secure cloud backup system that provides a secure layer for today's cloud storage service. It helps in eliminating of redundant data which is stored in backup. Yuan in [10] proposed a de-duplication system which is used to reduce the storage size of the tags for integrity check. This works to enhance the security of de-duplication and provide confidentiality. Bellar in [3] it transforms predictable message to unpredictable message to show how to protect data from unauthorized access. J. Li in [2] propose the technique to eliminate the redundant data instead of taking number of copies of same file. It also provide different encryption scheme that provide different security to popular and unpopular data. Another two-layered encryption scheme for high security with support of de-duplication is support for un popular data.

In convergent encryption R. D in [8] it offers keyless data security via dispersal algorithm. The algorithm uses the embedded random algorithm which breaks the data duplication in dispersal data. And ensure privacy in de-duplication. Bellare in [5] proposed new message locked encryption scheme which is useful in space-efficient secure outsourced storage.Xu.in [4] give a secured convergent encryption for encryption without considering issues of key-management and block-level de-duplication.

Some of the authors work on protocols proof of ownership for de-duplication system. K. Zhang in [14] they proposed two techniques proof of retrieve ability and proof of data possession to assure data integrity for cloud storage. And by using proof of ownership improves storage efficiency by securely removing unnecessary data which is stored on the cloud server. Proof of retrieve ability and proof of data possession are used in order to achieve both data integrity and storage efficiency which is the objective of proof of ownership. Halevi in [11] also use the concept of proof of ownership for de-duplication system, such that the client can prove to the cloud storage server that the client owns the file without uploading the file itself. One interesting concept in above techniques is that no one considers data privacy. Then Ng in [15] extends proof of ownership for encrypted files for security but do not address how to minimize the key management.

In some of the papers authors discuss about the twin clouds architecture. Bugiel in [7] provide and environment in which it consist of twin clouds for secure outsourcing of data. Zhang in [14] proposed the hybrid cloud technique to support more privacy then other proposed systems. The security model of our proposed model is similar to this related work, in related work they use the concept of private cloud but in our proposed system we use only public cloud.

III. PROPOSED ALGORITHM

In proposed system we make use of public cloud. A public cloud is based on the standard computing model. In this service providers' makes use of resources, like storage and applications, these applications and storage are available to the general public over the internet. Public cloud service are offered free or offered of a pay per usage model. Public cloud services are easy and inexpensive setup. There is no wastage of resources because you pay for what you use.

The proposed system proposed Cryptographic tuning which provides many benefits over the previous system, previous system makes use of convergent encryption which is insecure, since its short encryption key is generated from the input file in a deterministic way and could be leaked, roughly speaking convergent encryption is as insecure as

DOI: 10.15680/IJIRCCE.2015. 0309005

“hash-as- a-proof” method (i.e. using hash value hash as a proof of ownership of file in the presence of leakage). Therefore, all existing works on applying convergent encryption method to implement de-duplication of encrypted data are insecure in the bounded leakage setting of proof of ownership. In previous de-duplication system not support different authorization de-duplication check. This is important in many applications. In this authorized de-duplication check user issued a set of privileges during the system initialization. To solve the problem of de-duplication with different privileges in cloud system the proposed system consider cloud system, data owners outsource their data by utilizing public cloud and the data operations are managed in the same cloud. The de-duplication system support different de-duplication check in proposed under public cloud.

The proposed system is presented for carrying out secured authorized de-duplication process. It has many significant features. To increase the amount of information that can be stored on cloud by saving bandwidth and to eliminate duplicate copies of redundant data to preserve confidentiality of sensitive data while supporting de-duplication.

The proposed system address the problem of privacy preserving de-duplication in cloud computing and propose a new de-duplication system supporting for

Differential Authorization: Each authorized user is able to get his/her individual token of his file to perform duplicate check based on his privileges. Under this assumption, any user cannot generate a token for duplicate check out of his privileges or without the aid from the cloud server.

Authorized Duplicate Check: Authorized user is able to use his/her individual private keys to generate query for certain file and the privileges he/she owned with the help of public cloud, and public cloud performs duplicate check directly and tells the user if there is any duplicate.

A. *Description of the Proposed Algorithm:*

The proposed system contain implementation of authorized de-duplication system which contain following three model :

1. A Client program is used to model the data users to carry out the file upload process.
2. A Server program is used to model the public cloud which manages the private keys and handles the file token computation.
3. A Storage Server program is used to model the S-CSP which stores and de-duplicates files.

B. *The proposed Algorithm:*

The proposed system is based on the public cloud architecture. The private keys for privileges will not be issued to users directly, which will be kept and managed by the cloud server. In this way, the users cannot share these private keys of privileges in this proposed construction, which means that it can prevent the privilege key sharing among users in the above straightforward construction. To get a file token, the user needs to send a request to the private cloud server. The intuition of this construction can be described as follows. To perform the duplicate check for some file, the user needs to get the file token from the cloud server. The cloud server will also check the users’ identity before issuing the corresponding file token to the user. The authorized duplicate check for this file can be performed by the user with the public cloud before uploading this file. Based on the results of duplicate check, the user uploads this file.

Input: File F
Output: Encrypted file stored over cloud.

1. The object to be encrypted is validated to ensure it is suitable for this type of encryption. This generally means, at a minimum, the file is sufficiently long. (There is no point in encrypting say 3 bytes this way. Someone could trivially encrypt every 3-byte combination to create a reversing table.)
2. Some kind of hash of the decrypted data is created. Usually a specialized function just for this purpose is used, not a generic one like SHA-1. (For example, HMAC-SHA1 can be used with a specially-selected HMAC key not used for any other purpose.)
3. This hash is called the 'key'. The data is encrypted with the key (using any symmetric encryption function such as AES-CBC).

DOI: 10.15680/IJIRCCE.2015. 0309005

4. The encrypted data is then hashed (a standard hash function can be used for this purpose). This hash is called the 'locator'.
5. The client sends the locator to the server to store the data. If the server already has the data, it can increment the reference count if desired. If the server does not, the client uploads it. The client need not send the key to the server. (The server can validate the locator without knowing the key simply by checking the hash of the encrypted data.)
6. A client who needs access to this data stores the key and the locator. They send the locator to the server so the servers can look-up the data for them, and then they decrypt it with the key. This function is 100 percentages deterministic, so any clients encrypting the same data will generate the same key, locator, and encrypted data.

IV. Conclusion

The notion of authorized data de-duplication was proposed to protect the data security by including differential privileges of users in the duplicate check. We also presented new de-duplication check to supporting authorized duplicate-check tokens of files that are generated by the cloud server with private keys. The proposed system is secure from insider and outsider attacks. The proposed system implement new prototype model in which we used convergent encryption with modification version to deal with brute force attack using Cryptographic tuning to make better authorized de-duplication technique.

References

1. Li, Yan Kit Li, Xiaofeng Chen, Patrick P. C. Lee, Wenjing Lou. A Hybrid Cloud Approach for Secure Authorized De-duplication IEEE transaction VOL:PP NO:99 2014.
2. 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.
3. M. Bellare, S. Keelveedhi, and T. Ristenpart. Message-locked encryption and secure de-duplication. In EUROCRYPT, pages 296–312, 2013.
4. J. Xu, 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.
5. M. Bellare, S. Keelveedhi, and T. Ristenpart. Dupless: Serveraided encryption for de-duplicated storage. In USENIX Security Symposium, 2013.
6. P. Anderson and L. Zhang. Fast and secure laptop backups with encrypted de-duplication. In Proc. of USENIX LISA, 2010.
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. R. D. Pietro and A. Sorniotti. Boosting efficiency and security in proof of ownership for de-duplication. In H. Y. Youm and Y. Won, editors, ACM Symposium on Information, Computer and Communications Security, pages 81–82. ACM, 2012.
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. J. Yuan and S. Yu. Secure and constant cost public cloud storage auditing with de-duplication. IACR Cryptology ePrint Archive, 2013:149, 2013.
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. C. Ng and P. Lee. Revdedup: A reverse de-duplication storage system optimized for reads to latest backups. In Proc. of APSYS, Apr 2013.
13. 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, 201
14. K. Zhang, X. Zhou, Y. Chen, X.Wang, and Y. Ruan. Sedic: privacyawaredata 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.
15. W. K. Ng, Y. Wen, and H. Zhu. Private data de-duplication 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.

BIOGRAPHY

Ms. Naziya Tabassum is a student of M.TECH III semester in Computer Science and Engineering Department from Tulsiramji Gaikwad-patil College of Engineering and Technology, Wardha road, Nagpur (RTMNU). She is doing her project in under the guidance of **Prof. Roshani B. Talmale** HOD of CSE Department in Tulsiramji Gaikwad-Patil College of Engineering and Technology wardha road, Nagpur (RTMNU).