Volume-4, Issue-4, April-2015 • ISSN No 2277 - 8160

Research Paper Engineering

Two-step authentication with data de-duplication in Cloud

Manohar Vasantrao Rathod

ME (CSE) Student Shreeyash College of Engineering & Technology, Aurangabad.

ABSTRACT

Data optimization is a method for reducing the amount of storage space an organization and wants To protect its data in most companies the data storage system contains same copies of many parts of data for example some file may be appear in several different part by different users. De-duplication reduced this unwanted copies by saving only one copy of data and exchanging the other copies with reference that parts to first copy. Also in this paper we provide the multitier authentication scheme for user to protect user privacy against different types of attacks. Also we provide data privacy, integrity, authorization techniques for data storage in cloud.

KEYWORDS : Authentication, Authorization, Deduplication, authorized duplicate check, confidentiality, Integrity, hybrid cloud, convergent encryption.

INTRODUCTION

Cloud computing provides a low-cost, scalable, location-independent infrastructure for data management and storage. Owing to the population of cloud service and the increasing of data volume, more and more people pay attention to economize the capacity of cloud storage than before .Therefore how to utilize the cloud storage capacity well becomes important issue nowadays [20].

Types of Cloud Deployment Model.

1) Private cloud: - Private cloud is cloud infrastructure operated solely for a single organization, whether managed internally or by a third-party, and hosted either internally or externally.[2] Undertaking a private cloud project requires a significant level and degree of engagement to virtualized the business environment, and requires the organization to reevaluate decisions about existing resources.

2. Public cloud: - A cloud is called a "public cloud" when the services are rendered over a network that is open for public use. Public cloud services may be free.[6] Technically there may be little or no difference between public and private cloud architecture, however, security consideration may be substantially different for services (applications, storage, and other resources) that are made available by a service provider for a public audience and when communication is effected over a non-trusted network.

3. Hybrid Cloud: - Hybrid cloud is a composition of two or more clouds (private, community or public) that remain distinct entities but are bound together, offering the benefits of multiple deployment models. Hybrid cloud can also mean the ability to connect collocation, managed and/or dedicated services with cloud resources.

We proposed a scheme for hybrid cloud deployment model.

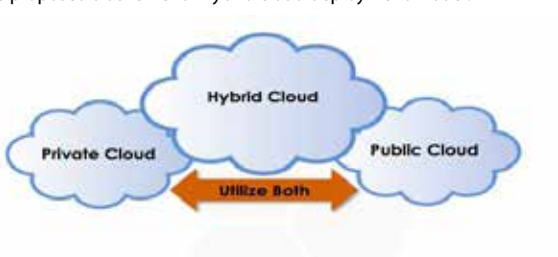

Fig 1.Strucutre of Hybrid Cloud

II.MOTIVATION

Cloud storage services are becoming very popular now a day. Cloud provides a better way of storage with efficient cost. One major problem with cloud is to manage huge amount of data. In order to manage data de-duplication technique is used. Although, de-duplication has many advantages but it has some security issues. This motivates us to propose a model which manage the security issues of de-duplication and provide authorized de-duplication in cloud

II. RELATED WORK

Data de-duplication is a technique for eliminating duplicate copies of data, and has been broadly used in cloud storage to reduce upload bandwidth and storage space. Predicting as it is, a coming up challenge to perform secure de-duplication in cloud storage. Although convergent encryption has been widely adopted for secure de-duplication, and it is a critical issue of making convergent encryption practical for reliably and efficiently manage a huge number of convergent keys. Our paper makes the first and best attempt to formally address the problem of achieving efficient and reliable key management in secure de-duplication [17]. We firstly introduce a standard methodology in which each and every user has their own separate master key for encrypting the convergent keys and outsourcing them to the cloud storage. However, we had a baseline or standard key management scheme which creates a huge number of keys as the users growing rapidly and requires users to enthusiastically protect the master keys. Storage efficiency functions such as compression and de-duplication afford storage providers" better utilization of their storage backend and the ability to serve more customers with the same infrastructure. Data de-duplication is the process by which a storage provider only stores a single copy of file owned by several of its users. There are four different de-duplication strategies, depending on whether de-duplication happens at the client side (i.e. before the upload happens) or at the server side, and whether de-duplication happens at a block level or at file level. De-duplication is most reinforcing when it is triggered at the client side, as it also saves upload bandwidth. For these reasons, de-duplication is a critical enabler for a number of popular and successful storage services that offer a cheap, remote storage to the broad public by performing client-side de-duplication, thus saving both the storage costs and network bandwidth. Well the data de-duplication is disputably one of the main reasons why the prices for cloud storage and cloud backup services have dropped so sharply. Unfortunately, de-duplication loses its effectiveness in conjunction with end-to-end encryption. End-to-end encryption in a storage system is the process by which data is encrypted at its source prior to ingress into the storage system. It is becoming an increasingly prominent requirement due to both the number of security incidents linked to leakage of unencrypted data and the tightening of sector-specific laws and regulations. Clearly, if semantically secure encryption is used, file de-duplication is impossible, as no one apart from the owner of the decryption key can decide whether two cipher texts correspond to the same plaintext.

PRAPOSED WORK

Data deduplication is one of the hottest technologies in storage right now because it enables companies to save a lot of money on storage costs to store the data and on the bandwidth costs to move the data when replicating it offsite for DR. This is great news for cloud providers, because if you store less, you need less hardware [4]. If you can de duplicate what you store, you can better utilize your existing storage space, which can save money by using what you have more effi-

ciently. If you store less, you also back up less, which again means less hardware and backup media. If you store less, you also send less data over the network in case of a disaster, which means you save money in hardware and network costs over time. The business benefits of data deduplication include:

A. Reduced hardware costs;
B.Reduced backup costs;
C.Reduced costs for business continuity / disaster recovery;
D.Increased storage efficiency; and
E.Increased network efficiency.

In the proposed system we are achieving the data deduplication by providing the proof of data by the data owner. This proof is used at the time of uploading of the file. 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 Paper ID: SUB15266 87 International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064 Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438 Volume 4 Issue 3, March 2015 www.ijsr.net Licensed Under Creative Commons Attribution CC BY 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. Also we provide user level authentication security techniques to access store data on cloud.

1. User Level Authentication

When user wants to access resources on the cloud, then user should login on to the cloud. Following are the steps to login on to the cloud.

1) User should enter Valid Email_Id and password in his login interface. User's system computes the secret key using stored values, which was already provided by the user at the time of registration.

2) The authentication server checks the user_id and password provided by the user with the user_id and password which was provided by the user at the time of registration.

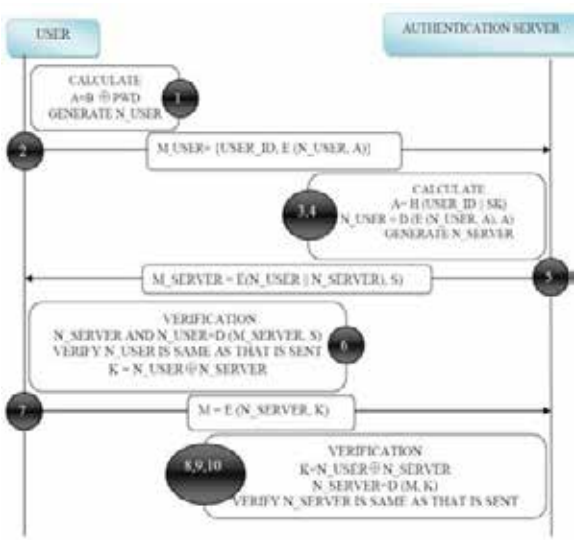

Fig 2. User Login Authentication.

3) After matching the user_Id and password authentication server generates the dynamic token from hash table and send it to the user's Email_Id for authentication.
4) User checks his Email for getting the dynamic token for further authentication.
5) User has to enter the token value for STEP-2 Authentication.
6) Authentication server matches the token with the dynamic token which was send by itself.
7) After matching the token authentication, user will authenticate and server provides access of resources to the user.

Password Change phase

This phase is used to provide facility of changing the Password. User has to provide his old password and new password to change his old password. Following are the steps.

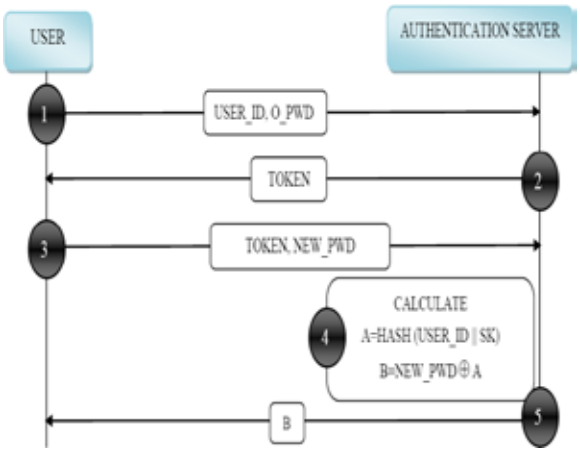

Fig 3.Password Change Phase

1) User has to provide his user_id and old password to change the password.
2) Authentication server checks the password with registered User_Id and password.
3) After the matching of User_Id and password it send the Dynamic token to the users Email_Id.
4) User has to provide token as well as new password to the Authentication server.
5) Authentication changes his old password to the new Password and sends the message to the user for change of Password.

2. Encryption/Decryption of files

Here we are using the common secret key k to encrypt as well as decrypt data. This will use to convert the plain text to cipher text and again cipher text to plain text. Here we have used three basic functions, KeyGenSE: k is the key generation algorithm that generates κ using security parameter 1. EncSE (k, M): C is the symmetric encryption algorithm that takes the secret κ and message M and then outputs the ciphertext C; DecSE (k, C): M is the symmetric decryption algorithm that takes the secret κ and ciphertext C and then outputs the original message M.

3. Confidential Encryption

It provides data confidentiality in deduplication. A user 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 [5].

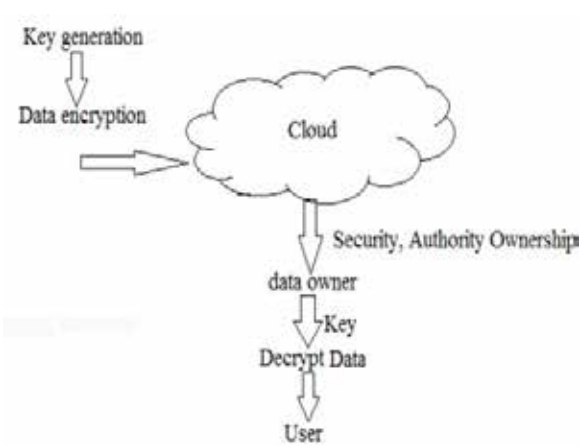

Fig 4. Confidential Data Encryption

4. Proof of Data

The users have to prove that the data which he wants to upload or download is its own data. That means he has to provide the convergent key and verifying data to prove his ownership at server. If there is a copy of this file and a match the privilege stored.

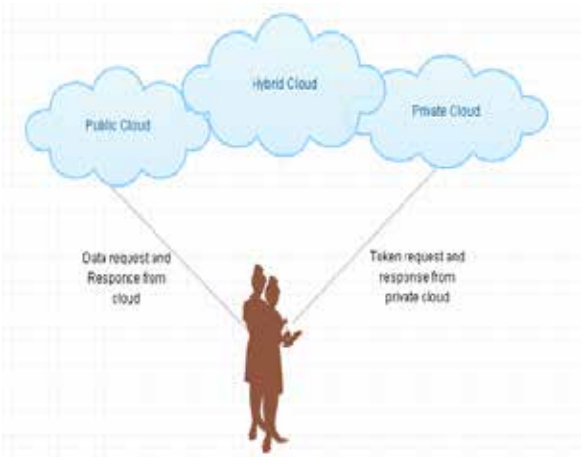

Fig 5.System Architecture

5. Deduplication Check

Data Deduplication is a specialized data compression technique for eliminating duplicate copies of repeating data. Can be categorized into two main strategies as follow, differentiated by the type of basic data Units [2].

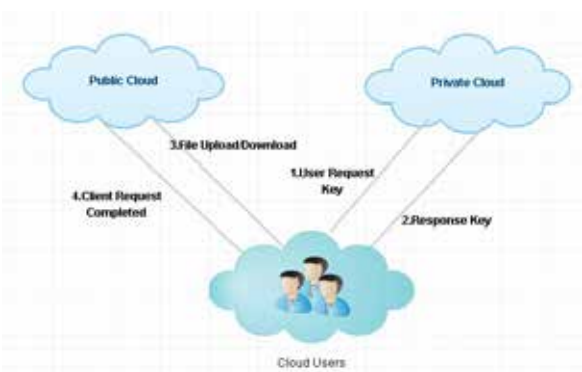

Fig 6.Authorized Deduplication Model

1).File-level deduplication: A file is a data unit when examining the data of duplication, and it typically uses the hash value of the file as its identifier. If two or more files have the same hash value, they are assumed to have the same contents and only one of these files will be stored.

2).Block-level deduplication: This strategy segments a file into several fixed-sized blocks or variable-sized blocks, and computes hash value for each block for examining the duplication blocks.

CONCLUSIONS

In this paper, we proposed an authorized data deduplication to protect the data security by including differential privileges of users in the duplicate check. We also presented several new hybrid deduplication 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. Also we used proofs of ownership protocol to ensure authorized user having access rights to data with user level authentication security.

REFERENCES

[1] M. Bellare, S. Keelveedhi, and T. Ristenpart. Dupless:Serveraided encryption for deduplicated storage. In USENIX Security Symposium, 2013. | [2] P. Anderson and L. Zhang. Fast and secure laptop backups with encrypted de-duplication. In Proc. Of USENIX LISA, 2010. | [3] J. Li, X. Chen, M. Li, J. Li, P. Lee, andW. Lou. Secure deduplication with efficient and reliable convergent key management. In IEEE Transactions on Parallel and Distributed Systems, 2013. | [4] 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. | [5] J. Li, X. Chen, M. Li, J. Li, P. Lee, andW. Lou. Secure deduplication with efficient and reliable convergent key management. In IEEE Transactions on Parallel and Distributed systems, 2013. | [6] C. Ng and P. Lee. Revdedup: A reverse deduplication storage system optimized for reads to latest backups. In Proc. of APSYS, Apr 2013. | [7] C.-K Huang, L.-F Chien, and Y.-J Oyang, "Relevant TermSuggestion in Interactive Web Search Based on ContextualInformation in Query Session Logs," J. Am.Soc. for Information science and Technology, vol. 54,no. 7, pp. 638-649, 2003. | [8] 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. | [9] 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. | [10] 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. | [11] S. Quinlan and S. Dorward. Venti: a new approach to archival storage. In Proc. USENIX FAST, Jan 2002. | [12] 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. | [13] 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. | [14] J. Stanek, A. Sorniotti, E. Androulaki, and L. Kencl. A secure data deduplication scheme for cloud storage. In Technical Report, 2013. | [15] Center Bo Wang, HongYu Xing "The Application of Cloud Computing in Education Informatization, Modern Educational Tech..." Computer Science and Service System (CSSS), 2011 International Conference on IEEE, 27-29 June 2011,978-1-4244-9762-1, pp 2673 – 2676 | [16] Mell P. and Grance T., "The NIST Definition of Cloud Computing", vol 53, issue 6, 2009. | [17] A Platform Computing Whitepaper, enterprise cloud computing: Transforming IT. Viewed 13 March 2010 [18] Dooley B 2010, Architecture requirement of The Hybrid Cloud''. Information Management Online, Viewed 10 February 2010. | [18] OpenSSL Project. http://www.openssl.org/. | [19] P. Anderson and L. Zhang. Fast and secure laptop backups with encrypted de-duplication. In Proc. of USENIX LISA, 2010. | [20] M. Bellare, S. Keelveedhi, and T. Ristenpart. Dupless: Server- aided encryption for deduplicated storage. In USENIX Security Symposium, 2013. |