Volume 4 Issue IV, April 2016
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Improving Security for De-Duplication System

Ms. Ranjana Phapale[1], Prof. B.S. Chunchure[2]
SPCOE College, Dept of Computer Engineering, Savitribai Phune University

Abstract— Removing duplicate data is nothing but Deduplication. De-duplication system is used in cloud computing to degrade storage space and the bandwidth of cloud. In this paper introduce some new algorithm that used for de duplication .We propose distributed systems with higher consistency in which the data chunks are sprade across many cloud servers. The security requirements of data are also achieved by introducing a deterministic secret sharing in distributed storage systems, instead of using convergent encryption as in previous systems. In this paper introduced new algorithms for de-duplication system. Cryptographic hash functions are generally used for the verification of data integrity. This paper provides an implementation of a newly selected cryptographic hash algorithm called Secure Hash Algorithm – Security analysis demonstrates that our de-duplication systems are secure in terms of the definitions specified in the proposed security model.
Keywords—Cloud, Data Storage, De-duplication, Security, Encoding, Decoding.

I. INTRODUCTION

Cloud computing is nothing but the use of resources (software and hardware) that delivered service over the network. The Cloud Computing name comes from the common use of a cloud shaped the complex infrastructure it contains in system diagrams. Cloud computing provide remote services with a user's data, software and hardware. Cloud computing consists of hardware and software resources made available on the Internet as managed emerging technologies services. These services typically provide access to software applications and high-end networks of server computers.

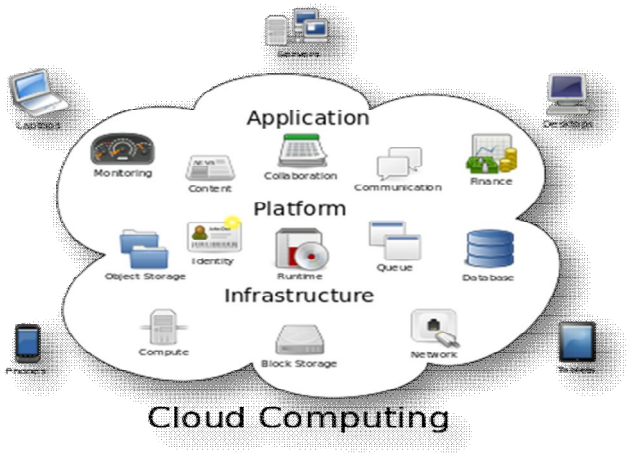

Figure 1.Shows Architecture of Cloud Computing

A. Benefits of cloud computing

1) *Economies of scale* – It increase productivity and volume output with less people. Your cost per unit, project or product decreases.
2) *Minimum personnel training* - It takes less people to do more work on a cloud, with a minimal learning requirement on hardware and software .
3) *Globalize work on the low rate* - People can access the cloud from anywhere, only they have an Interne connection.

a) *Reduce Costs* -There's no need to spend huge money on hardware, software or licensing fees.
b) *Improve accessibility* - You have access anytime, at anywhere, making your life easier.
c) *Reduce spending on technology* - Pay as you go. Maintain easy access to your information with minimal spending. Based on demand.

4) *Project Monitoring* - You can manage your project within budget and ahead of completion cycle times more effectively.

a) *Minimize licensing software or Hardware* - Stretch and grow without the need to buy expensive software licenses or programs.
b) *Streamline processes* - More work done in less time with less people.

II. RELATED WORK

A. Secure Data Deduplication

Authors: Mark W. Storer, Kevin Greenan, D.E.LongEthan L. Miller 2008.
They had developed a model for secure deduplicated storage: Authenticated model-file is itself encrypted using a unique key, sharing of this key is managed through the use of asymmetric key pair.
Limitation-This algorithm not increased the storage efficiency of system

B. Fast and secure laptop backups with encrypted deduplication

Authors: Anderson and Lang 2010.
This Paper introduced a algorithm and prototype That allow data to be encrypted independently without duplication and reduce the number of files
Limitation-This algorithm is appropriate for laptop.
Message-locked encryption and Secure deduplication.
Authors: M.Bellar, S.Keelveedhi and T.Ristenpart March 2013
Message-Locked Encryption (MLE),where the key under which encryption and decryption are performed is itself derived from the message. MLE provides a way to achieve secure deduplication
Limitation-No MLE scheme can achieve semantic security style privacy.

C. RevDedup: A Reverse Deduplication Storage System Optimized for Reads

To Latest Backups
Authors: Chun-Ho Ng and Patrick P. C. Lee June 2013
RevDedup removes duplicates from old data, thereby shifting fragmentation to old data while keeping the layout of new data as sequential as possible.
Limitation-RevDeduplicate technic apply only on fixed size chunk.

III. PRAPOSED SYSTEM

In Proposed System we have implementing deduplication system with improving reliability and proving security .For Duplicate or same copy checking we are using hash function which is calculated by SHA3 algorithm SHA is nothing but secure hash algorithm which is more powerful algorithm than previous hash key generation value algorithm, we are also proving a security with help of advance algorithm such as Advance Encryption. In this technique a file is dividing in to no of block or chunk and that chunk is also send to cloud to check duplicate or not at the end Encrypted file and hash
Key is store on to the cloud, for higher reliability file level and block De-duplication is used. The secret splitting technique is used for protect data. Our proposed structure supports both traditional De-duplication methods. Security, integrity and credibility can be achieved in our proposed system. In solution to kind of secret agreement attacks are considered.
Encloses tag consistency is another feature of our proposed system, can be derived .If the same value is stored in various cloud storage then De-duplication check by methods. Server establishes the collision attack that is not opposed. For higher reliability file level and block De-duplication is used. The secret splitting technique is used for protect data. Our proposed structure supports both traditional De-duplication methods. Security, integrity and credibility can be achieved in our proposed system. In solution to kind of secret agreement attacks are considered. In this security is concern with attack on data and security against server access .we can say that data is secure if limited number of access is provided to data access.

A. Work Flow for File Upload/Download

Volume 4 Issue IV, April 2016
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

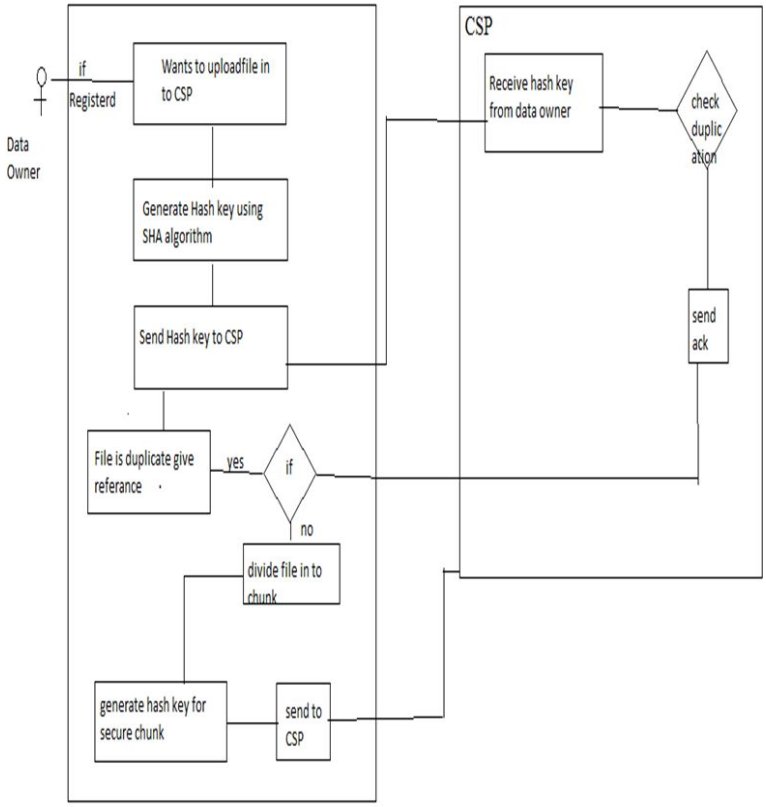

Figure 2: Workflow for File Upload /Download file

The above figure shows the architecture or flow of upload and download the file in De-duplication system, in this architecture we provide the overview of the system where user who want to upload the data or download the data from cloud storage server first user get authenticate if user is authorized then only he/she can upload or download the data from cloud storage once the user get authenticate then he upload or download the text file after verification of text file key generated by the SHA algorithm for duplication checking if file is not duplicate. Then only it will uploaded to the cloud or if the file is duplicate then cloud service provider gives the reference of already stored file.

IV. CONCLUSIONS

In this paper we are implementing the secure De-duplication systems to improve the security of client data. We proposed to support file-level and block-level data De-duplication. The security of tag consistency and integrity were achieved. We implemented our De-duplication systems using secret sharing scheme and demonstrated that it incurs small encryption/decryption overhead compared to the network transmission overhead in regular upload/download operations.

REFERENCES

[1] P. Anderson and L. Zhang" Fast and secure laptop backups with encrypted de-duplication, In Proc. of USENIX LISA, 2010.
[2] "Dupless: Server aided encryption for deduplicated storage. In USENIX Security Symposium, 2013.
[3] M. Bellare, S. Keelveedhi, and T. Ristenpart"Message-locked encryption and secure deduplication, In EUROCRYPT, pages 296 312, 2013.
[4] M. Bellare, C. Namprempre, and G. Neven" Security proofs for identity-based identication and signature schemes, J. Cryptology, 22(1):161, 2009.
[5] M. Bellare and A. Palacioschnorr identi_cation schemes: Proofs of securityagainst impersonation under active and concurrent attacks.
[6] In CRYPTO, pages 162177, 2002. S. Bugiel, S. Nurnberger, A. Sadeghi, and T. SchneiderTwin clouds: An archtecture for secure cloud computing. , InWorkshop on Cryptography and Security in Clouds (WCSC 2011), 2011.
[7] J. R. Douceur, A. Adya, W. J. Bolosky, D. Simon, and M. TheimerReclaiming space from duplicate _les in a serverless distributedle system. In ICDCS, pages 617624, 2002.
[8] 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
[9] J. Pepitone, “Dropbox’s password nightmare highlights cloud risks,” June 2011.
[10] 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.

[11] 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.

[12] 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.

[13] A. Yun, C. Shi, and Y. Kim, ''On Protecting Integrity and Confidentiality of Cryptographic File System for Outsourced Storage,'' in Proc. ACM CCSW, Nov. 2009, pp. 67-76.