Volume 5, Issue 10, October-2015 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Modern Secure Distributed Deduplication Systems with Improved Reliability

Gore Swapnali, Gore Supriya, Tengale Kanchan, Tengale Varsha, Asst.Prof. S. B. Bandgar
Prof. in Department of Computer engineering, S B Patil College of Engineering, Indapur, Dist– Pune, Savitaribai Phule Pune University, Maharashtra, India

Abstract*— For removing replication copies of data we use data deduplication process. As well as it is used in cloud storage to reduce memory space & upload bandwidth.only one copy for each file stored in cloud that can be used by number of users. Deduplication process helps to improve storage reliability. One more challenge of privacy for sensitive data also arises when they are outsourced by users to cloud. The aim of this paper is to make the first attempt formalize the idea of distributed reliable deduplication system. In our proposed system we are going to develop new distributed deduplication systems which is highly reliable. In deduplication process data chunks are distributed across multiple cloud servers. instead of using convergent encryption as in previous deduplication systems we use deterministic secret sharing scheme in distributed storage systems. So that we can achieve the required concepts for security that are data confidentiality and tag consistency. In the proposed security model, Security analysis demonstrates that our deduplication systems are secure.

Keywords*—Deduplication, reliability, distributed storage system, secret sharing, encryption

I. INTRODUCTION

In our proposed system we are going to use data deduplication process. First of all we must know what data deduplication is, it reduces the amount of data that needs to be physically stored by eliminating extra information and replacing after repetition of it with a pointer to the original. In data deduplication we remove unwanted copy of data and save the memory space.With the help of data deduplication method system reliability is improved as well as it avoid wastage of memory space. With the exquisite growth of digital data, deduplication techniques are widely used to backup data and minimize network and storage overhead by detecting and removing unwanted among data. Instead of keeping multiple data copies with the same content, deduplication eliminates unwanted data by keeping only one physical copy and referring other unwanted data to that copy.

Deduplication has received much attention from both academic and industry because it can more improves storage utilization and save storage space, especially for the applications with high deduplication ratio such as accession storage systems. For eliminating duplicate copies of data we use data deduplication technique. To reduce storage space and for uploading bandwidth mostly it has been used, In cloud storage. A various deduplication systems has been proposed based on number of policies such as client-side or server-side deduplications, file-level or block-level deduplications. The first attempt to describe the notion of distributed safe deduplication system. The main Aim of our proposed system is to describe the notion of distributed reliable deduplication system with more security.

We implement new distributed deduplication system, which has more reliability. In that data chunks are distributed across multiple cloud servers. Deduplication technique can save the memory space for the cloud storage service providers; it reduces the reliability of the system. Securityanalysis indicate that our deduplication systems are secure in terms of the definitions specified in this security model. As a proof of concept, we implement the proposed systems that indicate the acquired aerial is very limited in actual environments. Deduplication process improves storage utilization & it saves storage space .That's why it is useful in industry as well as in academic.It is useful in such application which has high deduplication ratio like as archival storage system. Furthermore, for the data privacy challenge is also arises more. The more sensitive data are redistributed by the users to cloud. Encoding have been usually Utilized, for to provide protection confidentiality before the redistributed data into cloud. Most commercial storage No of service providers are oppose to apply encryption over the data because it is impossible to make deduplication. The reason of that is the traditional encryption mechanism. In which including the public key encryption and symmetric key encryption have requirenumber of users to encrypt their data with own key. For the result of similar data copy of the number of users will lead to the different Data has been encrypted.To solve the problems of confidentiality and deduplication, for solving the problem of deduplication we implement notation of the convergent encryption.

II. LITERATURE SURVEY

Data deduplicationused for removing replication copies of data. Data deduplication techniques are very interesting techniques. The reliability produces stable and consistent results. They only focused on files without encryption, without considering the reliable deduplication over ciphertext. Ciphertext is also known as encrypted or encoded information. In

1997 M. Bellare introduced the idea of security and scheme for symmetric encryption in concentrate security framework [8].

They give different idea of security and analyze the good involution of reduction among them. They provide method of encryption using a block cipher, cipher block chaining and counter mode. Its have two goals .First is to study the idea of security for symmetrical encryption and second is to provide concrete security analysis of fixed symmetric encryption device. Convergent encryption provides data security in deduplication [10]. Bellare explains the message locked encryption system and give it's application in secure outsourced storage [8]. Encryption is used to achieve the data privacy. Li et al incline in block level deduplication having some key management problems, through several servers [10].

Bellare et al. displayed how to protect private data through the conversion of predictable message into the unpredictable message [8]. In their system, another third party knew the key server. It was introduced to produce the file tag to check the replicate copies. Stanek et al. accomplish better efficiency and security of outsourced data storage [9]. They provide differential security for all types of data.

D Harnik explains proof of ownership in Remote storage systems [5]. They identify attack that can lead to dropout data leakage and save time with the help of client side deduplication.

III. PROPOSED SYSTEM

To protect private data the secret sharing technique is used which is corresponding to distributed storage systems.In this paper the secret sharing technique is used for protection of private data.In detail a file is divide and encode into sections by using secret sharing technique. These sections will be distributed over many independant storage servers. A cryptanalysis hash value of the content will also be calculated and send to storage server as the mark of the fragment stored at each server. only the data user who first upload the data is required to calculate and distribute such secret shares and following users own same data copy do not need to calculate and stores these shares. Retrieve data copies owner must access a minimum number of storage server by a validation and obtain the secret shares to alter the data.In different way, the authorized uses will access the secret shares data copy. Another distinguishable feature of our proposal is that data completeness incloses tag consistency, can be derived. To explain further if the same value is stored in various cloud storage then deduplication check by methods.It cannot oppose the collision attack established by many servers. To our knowledge no related work on secure deduplication can rightly address, the reliability and tag consistency problem. The file level and block level deduplication is used for higher reliability.The secret splitting technique is used for protect data. Our proposed structure support both traditional deduplication methods.Privacy,credibility and integrity can be achieved in our proposed system.In solution to kind of secret aggrement attacks are considered.These are the attack on the data and the attack against servers. The data is secure when the opponent control limited number of storage servers.

Block Diagram/Architecture of Proposed System:-

When the user wants to upload and download the file from cloud storage at that time first user request to the web server for uploading file. It means only approved user can upload the file to web server for that purpose it use the proof of ownership algorithm . User to prove their relation of an owner to the thing possessed of data copies to the storage server. When file is uploaded it splits into blocks i.e by default size of block is 4KB.According to file size the block occurs. After that deduplication detection occurs.

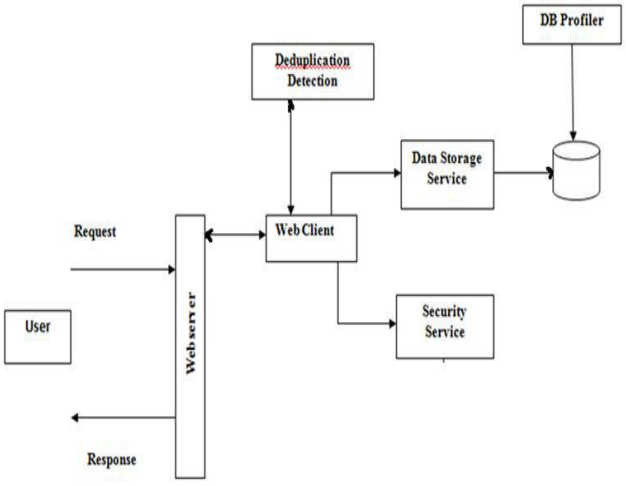

Fig: System Architecture

Web client having two services data storage service and security service . Data storage server contains all the uploaded files and Security service provide security to that files. DB profiler store all the metadata of the file

Workflow for File Upload /Download:-

Authorized user can access the file from cloud storage.

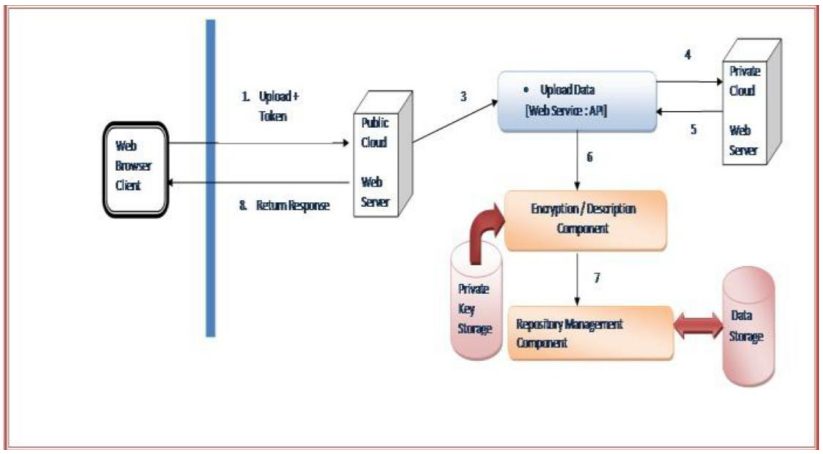

Fig: Workflow of upload/download

IV. MODULES

1. Secrete sharing scheme
2. The File-level Distributed Deduplication system
3. The Block-level Distributed Deduplication system

1. Secrete sharing scheme:-

In this module two algorithms are used which are Share and Recover. Share algorithm is used for partitioned and shared secret. With sufficient shares, Extracted and retrieved the secret with the help of Recover algorithm.

Share divides secret S into (k-r) fragments of same size , which produces r for random fragments of the equal size, and translate into simple language the k fragments using a non-systematic k-of–n erasion code into n shares of the similar size. Out of n shares the Recover adopts k from n shares as inputs. After that outputs the original secret S. A message authentication code (MAC) is a small section of knowledge used to authenticate a message and to provide integrity and authenticity certainty on the message. In our structure, the MAC is applied to derive the bonafides of the external sourced stored files.

2) File-Level Distributed Deduplication System:-

It support capable duplicate check, tags for each file will be calculated and send to storage cloud service provider. To prevent alignment invasion organized by the S-CSPs, tag collected at different storage servers.

System Setup: In our structure, the storage cloud service provider is considered to be n with identities denoted by id1, id2,…,idn respectively. To upload file F, the client communicate with S-CSPs to perform the elimination of duplicate data .For downloading file F, the client downloads the secret shares of the file from k out of storage servers.

3) Block-Level Deduplication System:-

In this part, we appear how to derive the fine grained block level distributed deduplication.In this system, the client also demands to perform the file level deduplication before uploading file. The user partition this files into blocks, if no duplication is found and performs block-level deduplicationsystem.The system set up is similar to file-level deduplication and also block size parameter will be defined.

V. CONCLUSION

We implement the secure distributed deduplication systems to improve the reliability of data while achieving the secret of the clients outsourced data.Four constructions were proposedto support file-level and fine-grained block-level data deduplication. The security of tag consistency and integrity were achieved.We implemented our deduplication systems using the Ramp secret sharing scheme and demonstrated that it incurs small encoding/decoding overhead compared to the network transmission overhead in regular upload/download operations.

REFERENCES

[1] J. S. Plank and L. Xu, "Optimizing Cauchy Reed-solomon Codes for fault-tolerant network storage applications," in NCA-06: 5th IEEE International Symposium on Network Computing Applications, Cambridge, MA, July 2006

[2] G. Ateniese, R. Burns, R. Curtmola, J. Herring, L. Kissner, Z. Peterson, and D. Song, "Provable data possession at Untrusted stores," in *Proceedings of the 14th ACM conference on Computer and communications security*, ser. CCS' 07. New York, NY, USA: ACM, 2007,

[3] A. Juels and B. S. Kaliski, Jr., "Pors: proofs of retrievability for large files," in Proceedings of the 14th ACM conference on Computer and communications security, ser. CCS '07. New York, NY, USA: ACM, 2007,

[4] M. W. Storer, K. Greenan, D. D. E. Long, and E. L. Miller, "Secure data deduplication," in *Proc. of StorageSS*, 2008.

[5] S. Halevi, D. Harnik, B. Pinkas, and A. Shulman-Peleg, "Proofs of ownership in remote storage systems." in ACM Conference on Computer and Communications Security, Y. Chen, G. Danezis, and V. Shmatikov, Eds. ACM, 2011, pp. 491–500.

[6] 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

[7] W. K. Ng, Y. Wen, and H. Zhu, "Private data deduplication protocols in cloud storage." in Proceedings of the 27th Annual ACM Symposium on Applied Computing, S. Ossowski and P. Lecca, Eds. ACM, 2012, pp. 441–446.

[8] M. Bellare, S. Keelveedhi, and T. Ristenpart, "Dupless: Serveraided encryption for deduplicated storage," in *USENIX Security Symposium*, 2013

[9] J. Stanek, A. Sorniotti, E. Androulaki, and L. Kencl, "A secure data deduplication scheme for cloud storage," in *Technical Report,* 2013.

[10] J. Li, X. Chen, M. Li, J. Li, P. Lee, and W. Lou, "Secure deduplication with efficient and reliable convergent key management," in *IEEE Transactions on Parallel and Distributed Systems*, 2014, pp. vol. 25(6), pp. 1615–1625.

