ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Reliability Secure Distributed Deduplication System

M.Aruna
M.Tech Student
Department of CSE
Prasad College of Engineering.

M.Srikanth
Associate Professor
Department of CSE
Prasad College of Engineering.

ABSTRACT:

The technology that eliminates duplicate and deduplication, cloud storage is widely used to reduce, but to fill the space. But there is only one copy of the file for each file there is the number of the foundation of the cloud, users can. Therefore, reducing the faith, better redundancy system for storage purposes. And when they could be removed from the protection of the privacy of the user sensitive data in the cloud. Consider health problems to solve on paper given to the first de- formalize the square distributed system can be trusted. We propose the new cart out of certain information, and divided it into multiple pieces, and they hand over the ratio of the square into the cloud. Requirements to use the information we should be saved and in secret, and, indeed, is a member of, and not in the before the meeting by the introduction of a participation in the mystery of the encryption depends deduplication system is a system of distributed system storage. Redundant reason for this is that the security of the appearance of the analysis showed a sufficiently secure a copy of the definition of health is set before us. This concept, he explains his position to carry out supreme and universal reason is limited.

INTRODUCTION

With the rapid growth of the techniques of de-duplication of digital data are widely active backup data security and reduce network and storage on the basis of the detection and elimination of redundancy information. Instead of placing multiple copies of the data in this content, duplication of sensitive data, keeping only one physical copy of other sensitive data to copy. Copy received great attention from academia and industry, as it can significantly improve storage utilization and to save storage space, especially in applications with high ratio of de-duplication as the storage system. The number of transcription system that offers various replication strategies such as client-side or server-of duplication, file level or block level of duplication. A brief description is given. Especially with the advent of cloud storage, technology de-duplication more accessible and enjoyable data to manage the growing volume of data storage services in the cloud for businesses and organizations to move to outsource data storage cloud providers of third parties, as evidenced by the many cases of real life. According to the research report IDC, the volume of data in the world will reach 40 trillion gigabytes by 2020 storage services business in the cloud as Dropbox, Google Drive and Mozy, using deduplication to save bandwidth and the storage cost by deduplicating customer -gefe. There are two types of duplication in terms of size: (i) Deduplication file level, which cuts between different files and delete these cuts in order to reduce the volume of needs, and (ii) of blocklevel Deduplication, which it is to identify and eliminate redundancies between data blocks. Files can be divided into small fixed size or variable block size. Use a fixed size blocks simplifies calculations of block boundaries,The blocks of variable size (for example, more than half of the fingerprint) provides better performance-duplication. Although the copy mode can save storage space in the cloud services, reduce reliability. data reliability is actuallyA very important issue of deduplication storage systems, because there is only one copy of each file is stored on a shared server by all owners. If a file / shared block away, a disproportionately large amount of data can be suitable for the unavailability of all files that share the same file / block. In case, if the block is measured in terms of the amount of file data that can be lost if a block is lost, the amount of data lost when a storage system block increments breaks with the number of each block. Therefore, how to ensure high reliability of

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

data deduplication system is a major problem. Most surgical deduplication system contains a single server configuration. However, as many of deduplication storage systems and cloud systems mean the lowest rate of users and applications, especially in the political system of storage if the data is variable and must be maintained in the long term. This requires that deduplication storage systems provide high reliability comparable to other systems available. In addition, the challenges of data privacy as well as a set of increasingly sensitive to external users in the cloud data. encryption mechanisms, as a rule, have been used to protect the confidentiality of data embrace outsourcing. service card is suitable for use in data encryption, as it enables deduplication.

The reason for this is that traditional encryption mechanisms, including encryption features key public-key encryption, users require different keys to encrypt data on your own. As a result, sensitive data copied to different users different cipher texts. Troubleshooting privacy and data deduplication, it has proposed the concept of convergent encryption to enforce data privacy and widely accepted, while performing deduplication. However, these systems achieve cost reduction of external data privacy caused the error. Therefore, how to protect the privacy and system reliability, while achieving deduplication storage cloud remains a challenge.

SYSTEM PRELIMINARIES

DATA OWNER

In this module, the owner of the data the data is loaded into the cloud server. For security reasons, the file owner encrypted data and then store the information in the cloud. The story can see a copy of the file with a deadly cloud server. Data owner can take a minimum of handling encrypted data files and data owner can see different cloud data, as well as a copy of a particular file. And you can create a flexible cloud servers use registered. And the owner of the option to move to the cloud, so you can move the files to the cloud server cloud server.

TRUSTEE

This module helps the connector to view a copy of the file if there is a cloud server, and you can see the many cloud servers. If it exists and then try downloading the file server to connect to the cloud automatically blocks access permission. If not, and you can upload the data to the owner of the file servers multi-cloud at a time.

CLOUD SERVER

Managed service provider cloud to cloud, to provide data storage services. Data encrypt data files and store them in the cloud to share the remote user. To access shared data files, data users download data files and design interest in the cloud, and then decipher.

REMOTE USER

This module is a flexible user connects using a username and password. After a secret file servers to cloud key, and to find the secret key. Since the secret key you are trying to download the file by entering the file name and the secret key to cloud server.

ATTACKER MODULE

In remote user module, while downloading time if remote user entered any wrong file name or secrete key then cloud servers treats him as attacker and moves his access permission to block/attacker list.

RELATED WORK

Technical data deduplication system Deduplication interesting techniques have been widely used in backup environments and enterprise data warehouse to reduce network overload detection and elimination of redundant data blocks. There are many schemes Deduplication offered by the scientific community. Deduplications address reliability. However, only focus on traditional file without encryption, regardless reliable deduplication cipher text. Li shows how to achieve a minimum of key duplication. However, they did not mention about the design file deduplication application support. Then show how to extend the road to building a strong user files deduplication. However, all these activities have not seen, and achieve the accuracy and completeness of the label. Convergent

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

encryption. convergent encryption ensures privacy of data deduplication. Bellare formalized this variable encryption lock loose and research programs and to provide a variable external storage space. There are several encryption options yield different convergent performance to ensure data deduplication. It is known that some of the storage vendors in the trade, such as the issue of encryption key management already a classic and convergent press Deduplication Block-level cloud and distribute these keys across multiple servers by encrypting the file . Bellare to show how to protect data privacy and unpredictable predicatable transform the message message. The system, called third-party server is the key to ensuring that the files to check for duplicate label. Stanek presented data encoding scheme that provides security on the contrary, people who want data. Sensitive data that is very rigid and traditional and conventional encryption. The encoding scheme and strong two-plySecurity while the support offered to those who want to dedupe data. Thus, they are better tradeoff between efficiency and security of the external data. The attacks, which can lead to loss of data in the storage system in the cloud-compatible client deduplication. In order to prevent these attacks, halevis proposed the concept of "clear property" (Pow) for the system deduplication, so that the client can effectively provide a storage server in the cloud that he / she maintains a file without having to load the file itself. A few more buildings PowMerkle hash tree offered support client-side deduplication, which includes dimension arrives configuration. Pietro Sorniotti offers genuine choice Pow scheme file prediction bit positions in a randomly selected as evidence file. Note that all of these systems do not consider data privacy. Recently, Xu presented Pow scheme allows eliminating duplicate in the customer arrives must establish random oracle model offered Pow design file, but not in the way of reducing overhead key management data. Of / PDP. Ateniese introduced the concept of protected data (PDP). The concept has been introduced to allow the cloud to customers, to ensure the integrity of the external data to the cloud in a cost effective manner. Juels proposed the concept of evidence recoverability (in). Compared to the PDP, the partner gets access to the cloud external evidence in the data server. This scheme Shacham and Waters is improved. The main difference between the two concepts is that by using error codes correction / deletion suffer an injury of the external data.

CONCLUSION

We have produced and distributed deduplication system to improve data reliability, while achieving the privacy of user data in external coding system. The four buildings has been suggested that support data file-level and block level deduplication fine grain. Achieved.We accuracy and system integrity Deduplication using the slope safety label distribution showed that it was caused by a coding / decoding of a lower load compared to the transmission grid load management / normal discharge.

REFERENCES

[1] Amazon, "Case Studies," https://aws.amazon.com/solutions/casestudies/#backup.

[2] J. Gantz and D. Reinsel, "The digital universe in 2020: Bigdata, bigger digi tal shadows, and biggest growth in thefar east," http://www.emc.com/collateral/analyst-reports/idcthe-digital-universe-in-2020.pdf, Dec 2012.

[3] M. O. Rabin, "Fingerprinting by random polynomials," Centerfor Research in Computing Technology, Harvard University, Tech.Rep. Tech. Report TR-CSE-03-01, 1981.

[4] J. R. Douceur, A. Adya, W. J. Bolosky, D. Simon, and M. Theimer,"Reclaiming space from duplicate files in a serverless distributedfile system." in ICDCS, 2002, pp. 617–624.

[5] M. Bellare, S. Keelveedhi, and T. Ristenpart, "Dupless: Serveraided encryption for deduplicated storage," in USENIX SecuritySymposium, 2013.

[6] ——, "Message-locked encryption and secure deduplication," inEUROCRYPT, 2013, pp. 296–312.

[7] G. R. Blakley and C. Meadows, "Security of ramp schemes," inAdvances in Cryptology: Proceedings of CRYPTO '84, ser. LectureNotes in Computer Science, G. R. Blakley and D. Chaum, Eds.Springer-Verlag Berlin/Heidelberg, 1985, vol. 196, pp. 242–268.

[8] A. D. Santis and B. Masucci, "Multiple ramp schemes," IEEETransactions on Information Theory, vol. 45, no. 5, pp. 1720–1728,Jul. 1999.

[9] M. O. Rabin, "Efficient dispersal of information for security, loadbalancing, and fault tolerance," Journal of the ACM, vol. 36, no. 2,pp. 335–348, Apr. 1989.

[10] A. Shamir, "How to share a secret," Commun. ACM, vol. 22, no. 11,pp. 612–613, 1979.

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

[12] S. Halevi, D. Harnik, B. Pinkas, and A. Shulman-Peleg, "Proofsof ownership in remote storage systems." in ACM Conference onComputer and Communications Security, Y. Chen, G. Danezis, andV. Shmatikov, Eds. ACM, 2011, pp. 491–500.

[13] J. S. Plank, S. Simmerman, and C. D. Schuman, "Jerasure: Alibrary in C/C++ facilitating erasure coding for storage applications- Version 1.2," University of Tennessee, Tech. Rep. CS-08-627,August 2008.

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

[15] C. Liu, Y. Gu, L. Sun, B. Yan, and D. Wang, "R-admad: Highreliability provision for large-scale de-duplication archival storagesystems," in Proceedings of the 23rd international conference onSupercomputing, pp. 370–379.

[16] M. Li, C. Qin, P. P. C. Lee, and J. Li, "Convergent dispersal:Toward storage-efficient security in a cloud-of-clouds," in The 6thUSENIX Workshop on Hot Topics in Storage and File Systems, 2014.

[17] P. Anderson and L. Zhang, "Fast and secure laptop backups withencrypted de-duplication," in Proc. of USENIX LISA, 2010.

[18] Z. Wilcox-O'Hearn and B. Warner, "Tahoe: the least-authorityfilesystem," in Proc. of ACM StorageSS, 2008.

[19] 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 andversion control," in 3rd International Workshop on Security in CloudComputing, 2011.

[20] M. W. Storer, K. Greenan, D. D. E. Long, and E. L. Miller, "Securedata deduplication," in Proc. of StorageSS, 2008.