RESEARCH ARTICLE OPEN ACCESS

A Survey paper on Secure Data Deduplication with OTP System in Cloud

Harsh Gupta [1], Ashvini Vairagade[2] , Prof. Shweta Bhelonde[3]
Computer engineering,S.R.P.C.E.,Nagpur,India,
Computer engineering,S.R.P.C.E.,Nagpur,India,
Computer engineering,S.R.P.C.E.,Nagpur,India,

Abstract :
Cloud computing offers a new way of service provision by rearrange various resources over the Internet. The most important and trendy cloud service is data storage. In order to maintain the privacy of data holders, data are often stored in cloud in an encrypted form. However, data introduce new concept of data to enhanced challenge of all over solutions of data deduplication suffer from security problem. They cannot flexibly support data access control and revocation. Therefore, few of them can be readily deploy in practice. In this paper, we proposed the security awareness and the flexibly searching of the large amount of data to help the distributing the object and doesn't exist deduplicate data. It integrates cloud data deduplication with access control. We calculate its performance based on the scheme for potential useful deployment, particularly for large amount of data deduplication in cloud storage.

INTRODUCTION:

We need to store our data with the security purpose any type of data we can save but the backup problem is very major problem to everyone and doesn't handle with proper manner is so difficult. As an example of the IT sectors performing various numbers of important data but main problem storage capacity and the security problem also backup recovery. It create the critical situation for company and the disadvantage of wasting time and space without security, but cloud provide storage space with security awareness[1][5]The benefits of storing data to the cloud for sharing in sequence among source to destination simplify moving data between different devices, and for small businesses to back up and provide backup recovery capabilities[1] If you intend to move large amounts of data over a network and the flexibility of the internet services, you need to be cognizant of network bandwidth requirements, data security and the total costs of providing those services to end users, when providing services for data storage and protection [2][3] this method being used to accomplish data deduplication across multiple clients, where the costs to provide of multiple type of storage. There is many technique to deduplicate records, so service provider and their clients need to be aware of the difference between the available solutions and the collision they may have on protect and the ability to capably shift, defend and store data in the cloud in a efficient manner. Data deduplication is one of the most new storage right now since, if you can deduplicate which you store, you can better consume your presented storage space, which can save time

more efficiently. [4] If you also send more less data over the network in case of a breakdown, which means you save money in hardware and network costs over time [5][17].

REVIEW OF LITERATURE

Sr.no	Paper name	year	Author name	Description
1	SAFE: Structure-Aware File and Email Deduplication for Cloud-based Storage Systems	2013	Daehee Kim, Sejun Song, Baek-Young Choi	intend a novel Structure-Aware File and Email deduplication (SAFE) plan that achieves both well-organized and efficient data deduplication at a source site for cloud-based storage
2	Reducing fragmentation impact with forward knowledge in backup systems with deduplication	2015	Michal Kaczmarczyk, CezaryDubnicki	Deduplication of backups is very effective in reduction storage, but may also cause significant restore reduce speed. This trouble is caused by data fragmentation
3	Deduplication on Encrypted Big data in using HDFS Framework	2017	Apurva Nandurkar, Mrunalineepatole	Data de-duplication is a particular data compression method for eliminating duplicate copies of repeat data in storage.
4	Secure Data Deduplication in Cloud	2017	Arpitha .R, Pavithra	Cloud provider offer potentially never-ending storage space, where users can use as much space as they can and vendors continually look for techniques which aimed to minimize not needed data (multiple copies) and make best use of space savings.
5	A survey and classification of storage deduplication system	2014	Joao Paulo, Jose Pereira	The usual removal of duplicate data has been used for a long time in archival and back up system.
6	Proxy re-encryption (pre) based deduplication schema on encryption big data for cloud	2017	A.ShyamalaDevi, Dr.V.SaiShanmuga Raja	Security and assess the presentation of the propose scheme through analysis and imitation. The result shows its efficiency, effectiveness and applicability.
7	Data Hiding System Using Cryptography & Steganography : A	2015	Aarti Mehndiratta	Cryptography is a widely used technique that encrypts plain text to generate cipher (encrypted) text. Data that can be read and understood without any special measures is called plaintext or clear text.

	Comprehensive Modern Investigation			
8	DE-DUPLICATION OF DATA IN CLOUD	2016	R. SHOBANA*, K. SHANTHA SHALINI, S. LEELAVATHY and V. SRIDEVI	Cloud computing is one of the emerging technology, which helped several organizations to save money and time adding convenience to the end users. Thus the scope of cloud storage is vast because the organizations can virtually store their data's without bothering the entire mechanism. Cloud Computing provides key advantage to the end users like cost savings, Able to access the data irrespective of location, performance and security.
9	A Survey on Secure Deduplication of Data in Cloud Storage	2015	Babaso D. Aldar, Vidyullata Devmane	a. File-level de-duplication b. b. Block-level de-duplication
10	Study on secure data duplication system with application awareness over cloud storage systems	2015	Dipti Bansod, Amar buchade	The deduplication storage system deduplicates data with application awareness index structure. This system consists of two major components, a front-end deduplication and a cloud storage system as back-end.

METHODOLOGY

Data Deduplication works by compares objects (usually index files or unique blocks) and removes objects (copies) that already exist in the data set they save space and time.

In Data Deduplication method we divide the input data into unique blocks and a hash value is calculated for each of every block and uses the searching result of source to destination. Then using these hash values we can determine whether another block of same data has already been stored. If a comparable data file is found then restore the duplicate data with a reference to the object

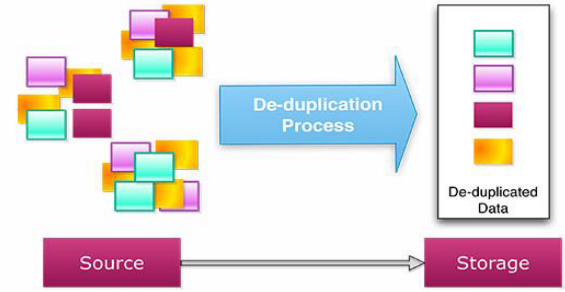

Fig. search deduplicate data

In simplified terms, data Deduplication compares substance (usually files or blocks) and removes substance (copies) that already exist in the data set. Simply place, the process consists of four steps:

1. Divide the input data into blocks

2. Calculate a hash value for each block or chunks of data.

Hx (Z1) =Hx (Z2)⟺ Y1=Y2

Hy (Z2) =Hy (Z2)⟺ Y1=Y2

A (K1, E (K2, Y)) =Y⟺ K1=K2

3. Use these values to determine if another block of the same data has already been stored.

4. Replace the duplicate data with a reference to the object already in the database.

one time the data is break, an index can be created from the results, and the duplicates can be found and eliminated. Only a single instance of every large piece is stored. The actual process of data Deduplication can be implements in a number of different ways. You can reduce duplicate data by simply comparing two files, object file and so many data we can store here.

In this method, stored the large amount of data to the source and the chunks the file depend their size of the file and documents. Proxy re-encryption connects with the server and merged the connectivity with the internet services via dropbox application to providing the storage capacity with the needed space and another method provide the encryption key refer to the deduplication data and the security factor with the flexibility. Storage time with encrypted data and the hash key value send source to the destination.

SUMMARY

This paper notion the information about data deduplication for the dropbox cloud based systems. It includes the methods that are used to achieve cost useful storage and efficient bandwidth usage by deduplication. The core concept involve eliminate the duplicate copies of the repeated data by using hashing algorithms. However, data deduplication is the most critical element for scivilizing efficiency of the cloud system. This technique will play a major role in the cloud based services for storing backup data by both medium and large enterprises with the security of OTP concept

The thought of official data de-duplication technique is listening carefully data compression method which eliminates unneeded data as well as improves storage and bandwidth operation. which encrypt data before outsourcing. safety analysis demonstrates that the schemes are protected in terms of insider and outsider attacks.

REFERENCES:

[1] M. Bellare, S. Keelveedhi, and T. Ristenpart, “DupLESS: Server aided encryption for deduplicated storage,” in Proc. 22nd USENIX Conf. Secure., 2013, pp. 179–194.

[2] J. R. Douceur, A. Adya, W. J. Bolosky, D. Simon, and M. Theimer, “Reclaiming space from duplicate files in a serverless distributed file system,” in Proc. IEEE Int. Conf. Distrib. Compute. Syst., 2002, pp. 617–624, doi:10.1109/ICDCS.2002.1022312.

[3] G. Wallace, et al., “Characteristics of backup workloads in production systems,” in Proc. USENIX Conf. File Storage Technol., 2012, pp. 1–16.

[4] Z. O. Wilcox, “Convergent encryption reconsidered,” 2011. [Online]. Available: http://www.mailarchive.com/cryptography@ metzdowd.com/msg08949.html

[5] G. Ateniese, K. Fu, M. Green, and S. Hohenberger, "Improved proxy re-encryption schemes with applications to secure distributed storage," ACM Trans. Inform. Syst. Secure., vol. 9, no. 1, pp. 1– 30, 2006, doi:10.1145/1127345.1127346.

[6] Opendedup. (2016). [Online]. Available: http://opendedup.org/ [10] D. T. Meyer and W. J Bolosky, "A study of practical deduplication," ACM Trans. Storage, vol. 7, no. 4, pp. 1–20, 2012, doi:10.1145/2078861.2078864.

[7] J. Pettitt, "Hash of plaintext as key?" (2016). [Online]. Available: http://cypherpunks.venona.com/date/1996/02/msg02013.html

[8] The Freenet Project, Freenet. (2016). [Online]. Available: https:// freenetproject.org/

[9] M. Bellare, S. Keelveedhi, and T. Ristenpart, "Message-locked encryption and secure deduplication," in Proc. Cryptology—EUROCRYPT, 2013, pp. 296–312, doi:10.1007/978-3-642-38348-9_18.

[10] D. Perttula, B. Warner, and Z. Wilcox-O'Hearn, "Attacks on convergent encryption." (2016). [Online]. Available: http://bit.ly/ yQxyvl

[11] C. Y. Liu, X. J. Liu, and L. Wan, "Policy-based deduplication in secure cloud storage," in Proc. Trustworthy Comput. Serv., 2013, pp. 250–262, doi:10.1007/978-3-642-35795-4_32.

[12] P. Puzio, R. Molva, M. Onen, and S. Loureiro, "ClouDedup: Secure deduplication with encrypted data for cloud storage," in Proc. IEEE Int. Cof. Cloud Comput. Technol. Sci., 2013, pp. 363–370, doi:10.1109/CloudCom.2013.54.

[13] Z. Sun, J. Shen, and J. M. Yong, "DeDu: Building a deduplication storage system over cloud computing," in Proc. IEEE Int. Conf. Comput. Supported Cooperative Work Des., 2011, pp. 348–355, doi:10.1109/CSCWD.2011.5960097.

[14] Z. C. Wen, J. M. Luo, H. J. Chen, J. X. Meng, X. Li, and J. Li, "A verifiable data deduplication scheme in cloud computing," in Proc. Int. Conf. Intell. Netw. Collaborative Syst., 2014, pp. 85–90, doi:10.1109/INCoS.2014.111.

[16] J. Li, Y. K. Li, X. F. Chen, P. P. C. Lee, and W. J. Lou, "A hybrid cloud approach for secure authorized deduplication," IEEE Trans. Parallel Distrib. Syst., vol. 26, no. 5, pp. 1206–1216, May 2015, doi:10.1109/TPDS.2014.2318320.

[17] P. Meye, P. Raipin, F. Tronel, and E. Anceaume, "A secure twophase data deduplication scheme," in Proc. HPCC/CSS/ICESS, 2014, pp. 802–809, doi:10.1109/HPCC.2014.134.

[18] J. Paulo and J. Pereira, "A survey and classification of storage deduplication systems," ACM Comput. Surveys, vol. 47, no. 1, pp. 1–30, 2014, doi:10.1109/HPCC.2014.134.

[19] Y.-K. Li, M. Xu, C.-H. Ng, and P. P. C. Lee, "Efficient hybrid inline and out-of-line deduplication for backup storage," ACM Trans. Storage, vol. 11, no. 1, pp. 2:1-2:21, 2014, doi:10.1145/2641572.

[20] M. Fu, et al., "Accelerating restore and garbage collection in deduplication- based

backup systems via exploiting historical information," in Proc. USENIX Annu. Tech. Conf., 2014, pp. 181–192.

[21] M. Kaczmarczyk, M. Barczynski, W. Kilian, and C. Dubnicki, "Reducing impact of data fragmentation caused by in-line deduplication," in Proc. 5th Annu. Int. Syst. Storage Conf., 2012, pp. 15:1–15:12, doi:10.1145/2367589.2367600.

[22] M. Lillibridge, K. Eshghi, and D. Bhagwat, "Improving restore speed for backup systems that use inline chunk-based deduplication," in Proc.USENIX Conf. File Storage Technol., 2013, pp. 183–198. [19] Fahlman, Scott E. and Lebiere, Christian: The Cascade- Correlation Learning Architecture, Neural Information Processing Systems 2, page 524-532, 1990