A Survey on Cloud Storage Techniques

[1]D.Mohanapriya, [2]V.Rajkumar
[1]Assistant Professor (CSE), [2]Student
Nandha College of Technology, Erode

***Abstract* - Cloud computing is which describing the new computing paradigm in the healthcare sector that is besides other business domains. Huge numbers of health organizations have started that are shifting the electronic health information that describes the cloud environment. That Introducing the cloud services in the health sector that are not only facilitates the exchange of electronic medical records among the hospitals and clinics, but also enables the cloud that are act as a medical record storage center. In this paper, we describe the existing results of coding techniques for cloudstorage techniques. Specifically, we present these coding techniques that are splitted into two categories: first one is regenerating codes and then locally repairable codes. These two codes meet the requirements at where cloud storage along two different axes: optimizing bandwidth and I/O overhead. We present an overview that describes the recent advances in these two categoriesof coding techniques. Moreover, we introduce the major ideas of some specific coding techniques.That are high level and that are discussing about the motivations and performance.**

***Key words* - erasure coding; cloud storage; regenerating codes; locally repairable codes**

I. INTRODUCTION

The decreasing rate of bandwidth and storage price are technique, that has been observed and that are helps to the IT companies such as Google, Microsoft, Amazon and etc, That are uses to build their services inside data centers which help to provide the services globally through a high-bandwidth network. This new paradigm providing the computing services .This context of cloud computing, the storage has not an important component of large-scale cloud services. That are also intend to provide a virtual storage infrastructure in a pay-as-you-go manner, such as Amazon . Meanwhile this volume of data stored inside data centers that are observe to be grow which are much faster than Moore's Law. This technique reported about the storage space which used for photo.This has been over 20 PB in 2011 and that is still increasing by 60 TB every week.

To meet their requirements of the massive volume of the storage, the cloud storage system has to scale out from the operation, , storing data in a large number of commodity disks is the example for this operation. In this process, it describe the major challenge for cloud storage systems that are use to maintain the data integrity, that happen due to both the large no of disks and their commodity nature. Even though the disk failures at number of times in a small portion at where the data are placed inside, there still have chance to be a large number of such failures happen everyday due to the large number of disks. The facebook cluster having 3000 nodes.There are typically 20 repairs are triggered on everyday. The storage devices and the other systems in the data center, such as the networking and power systems, may cause outages in the data center which making the data unavailable or even get lost

II. II. CLOUD STORAGE TECHNIQUES

The operations of Cloud storage techniques has recently grown in advance.The popular methods as mentioned above were analysed in detail.

A. Naive Method

The idea behind this technique is to comparisons the data. In this technique, the client will compute the different hash value for the file F and having key K and that are subsequently will send the file F to the server. Clients are having different connection of keys and hash values. So it helps to perform multiple check on the file F. Whenever client wants to check the file it release some specific key K and that sends it to the server, which asked after the recomputed of hash value, that are based on file F and key K . Now server would reply back to the client at where the client has the hash value for comparison. This method gives the proof that helps the server to find the original file F. This method has high overhead at every time hashing process is run over the entire file which having very high computation cost.

B. Original Provable Data Possession

In this technique, the data is based on pre-processed technique that describes the operation at where the data's are sending before to the cloud server. Here the data is fulfilled with tag value. Now the entire data has been sent over to the server and that are perform at the client side meta-data. This meta-data perform the operation of verification. This happen when user need for it.Suppose if user wants to check the integrity, it will send the challenge to the server when the time server will respond with the data. The client will compare the reply data with a local meta-data. This technique describes the operation where client will say whether the data is modified or not. Original PDP has low computation and storage is overhead. It supports both encrypted data which are at plain data format. It offers public verifiability operations. It increases the efficient because small portion of the file that is needs to be accessed to generate the proof on the server. This technique is only one which applicable to the static files (i.e.

append-files only). Homomorphic hashing technique is the one which used to compose multiple block inputs into a single value . that are helps to reduce the size of proof.

C. Information Flow Graph

To investigate about bandwidth consumption that repairs the storage with erasure coding, Dimakis et al. proposed which uses the information flow graph. This tool used at analysis of network coding. That are acted as a model which characterize the tradeoff between storage and bandwidth..That describe the complete information about information flow graph. All servers are categorized into source, storage nodes, and the data collector. The source are the one which denotes the server, that the data object is originated. Suppose that the rate of the data object is M bit, then the process is encoding. The coded blocks of bits are disseminated which splitted into n storage nodes. Particularly, the source is the one which represent by a vertex and the storage node. That are described by two vertices in the information flow graph. The weight of the edges are correspond to the amount of data stored in the storage node. After the dissemination process, all n storage nodes store bits and the key are sufficient to recover the original data object,that suggesting the $k > M$. A virtual vertex is the one which called as data collector. That are able to connect any k storage nodes which helps to recover the original data object. When the storage node are fails, a newcomer does not exist. That are just connect to k available storage nodes, but d storage nodes are normally act as providers ($d > k$).

D. Exact Regenerating Codes

The interference alignment is the important tool that are used to construct the exact regenerating codes , which initially proposed for wireless communication techniques. The basic idea behind the interference alignment is that the undesired vectors, that can be eliminated by aligning them onto the same linear subspace. It describe how interference alignment helps to achieve exact regenerating codes. We describe the data that are encoded with (n=4, k=2, d=3) MSR codes, In the each storage node, the coded block contains atleast two coded segments, such as A1, A2. That are present at the failed storage node. To perform the operation of recover, A1 and A2 placed at the newcomer. That are contacts the 3 storage nodes which are denoted by providers and this downloads the half of a coded block. Notice that each provider having the coded segments, that are containing the components of B1 and B2, which are undesirable to the newcomer. To eliminate the status at B1 and B2, each provider should send a segment in which B1 and B2 are aligned at same linear subspace of B1 C B2. Clearly, B1 C B2 describe the eliminate operation which are unknown and A1 and A2 would be decoded by solving three equations which three are unknowns. The process of extraction at MBR codes, Rashmi et al. proposed a Product-Matrix construction which helps to construct explicitly. The Product-Matrix helps to produce the vector MBR codes such that a coded block denotes the multiple coded segments.

III. CONCLUSION

In this survey various techniques of cloud storage were described in detail.These techniques are most important which uses to describe the various virtualized i/o performance of cloud storage services not comparable to local disk yet. Academic cloud systems are not providing a rich set of storage services so far .Performance tests for commercial storage services in future.More investigation on design and implementation details Include emerging services from other providers.

REFERENCES

[1] Amazon.com, “Amazon s3 availablity event: July 20, 2008”, Online at http://status.aws.amazon.com/s3-20080720.html, 2008.

[2] P. S. Browne, “Data privacy and integrity: an overview”, In Proceeding of SIGFIDET ’71 Proceedings of the ACM SIGFIDET (now SIGMOD), 1971.

[3] A. Cavoukian, “Privacy in clouds”, Identity in the Information Society, Dec 2008.

[4] R. Gellman, “Privacy in the clouds: Risks to privacy and confidentiality from cloud computing”, Prepared for the World Privacy Forum, online at http://www.worldprivacyforum.org/pdf/WPF Cloud Privacy Report.pdf,Feb 2009.

[5] N. Gruschka, M. Jensen, “Attack surfaces: A taxonomy for attacks on cloud services”, Cloud Computing (CLOUD), 2010 IEEE 3rd International Conference on, 5-10 July 2010.

[6] W. Itani, A. Kayssi, A. Chehab, “Privacy as a Service: Privacy-Aware Data Storage and Processing in Cloud Computing Architectures,” Eighth IEEE International Conference on Dependable, Autonomic and Secure Computing, Dec 2009

[7] M. Jensen, J. Schwenk, N. Gruschka, L.L. Iacono, “On Technical Security Issues in Cloud Computing”, IEEE International Conference on Cloud Computing, (CLOUD II 2009), Banglore, India, September 2009, 109-116

[8] A. Juels and B. S. Kaliski, PORs: Proofs of retrievability for large files, ACM CCS, (2007) 584-597.

[9] G. Ateniese, R. D. Pietro, L. V. Mancini, and G. Tsudik, Scalable and efficient provable data possession, SecureComm, 2008.

[10] C. Erway, A. K ¨ upc¸ ¨u, C. Papamanthou, and R. Tamassia, Dynamic provable data possession, Proc. 16th ACM conference on Computer and communications security, (2009) 213-222.

[11] K.D. Bowers, A. Juels, and A. Oprea, HAIL: A high-availability and integrity layer for cloud storage, Proc. 16th ACM conference on Computer and communications security, (2009) 187-198.

[12] S. Eswaran, Dr. S. Abburu, Identifying Data Integirty in the Cloud Storage, International Journal Of Computer Science Issues, Vol.9 Issue 2, (2012), ISSN 1694-0814

[13] R. Sravan and Saxena ,”Data integrity proofs in cloud storage” in IEEE 2011.

[14] R. Pandya, K. Sutaria, “An analysis of privacy techniques for data integrity in the cloud environment”, International Journal of Computer and Electronics Engineering,(Dec 2012) ISSN: 0975-4202