

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 5, May 2019

A Survey on Data Sharing in Dynamic Group Using Store SIM

P. M. Bisaguppi, D. B. Kshirsagar

M. E Student, Department of Computer Engineering, Sanjivani College of Engineering, Kopergaon, Maharashtra, India

Professor, Department of Computer Engineering, Sanjivani College of Engineering, Kopergaon, Maharashtra, India

ABSTRACT: Many schemes have been recently advanced for storing data on multiple clouds. Distributing data over different cloud storage providers (CSPs) automatically provides users with a certain degree of information leakage control, for no single point of attack can leak all the information. However, unplanned distribution of data chunks can lead to high information disclosure even while using multiple clouds. In this paper, we study an important information leakage problem caused by unplanned data distribution in multi cloud storage services. Then, we present **Store Sim**, an information leakage aware storage system in multi cloud. Store Sim aims to store syntactically similar data on the same cloud, thus minimizing the user's information leakage across multiple clouds. Our design an approximate algorithm to efficiently generate similarity-preserving signatures for data chunks based on Min Hash using hashing Methods and also designs a function to compute the information leakage based on these signatures. Next, we present an effective storage plan generation algorithm based on clustering for distributing data chunks with minimal information leakage across multiple clouds. New Methods are used Block Level whenever data are Clustering for distributing data chunks it will put into randomly blocks.

KEYWORDS: Multi cloud storage, information leakage, system attack-ability, remote synchronization, distribution and optimization

I. INTRODUCTION

With the increasingly number of devices such as laptops, cell phones and tablets, users require ubiquitous and massive network storage to handle their ever-growing digital lives. To meet these demands, much cloud-based storage and file sharing services such as Drop box, Google Drive and Amazon S3 have gained popularity due to the easy-to-use interface and low storage cost. However, these centralized cloud storage services are criticized for grabbing the control of users' data, which allows storage providers to run analytics for marketing and advertising One possible solution to reduce the risk of information leakage is to employ multi cloud storage systems in which no single point of attack can leak all the informaion. A malicious entity, such as the one revealed in recent attacks on privacy would be required to coerce all the different Cloud Server Provider on which a user might place her data, in order to get a complete picture of her data. Put simply, as the saying goes, do not put all the eggs in one basket. Yet, the situation is not so simple. CSPs such as Dropbox, among many others, employ rsync-like protocols to synchronize the local file to remote file in their centralized clouds. Every local file is partitioned into small chunks and these chunks are hashed with fingerprinting algorithms such as SHA1, MD5 Thus, a file's contents can be uniquely identified by this list of hashes. For each update of local file, only chunks with changed hashes will be uploaded to the cloud. This synchronization based on hashes is different from like as protocols that are based on comparing two versions of the same file line by line and can detect the exact updates and only upload these updates in a patch style. Instead, the hash based synchronization model needs to upload the whole chunks with changed hashes to the cloud. Thus, in the multi cloud environment, two chunks differing only very slightly can be distributed to two different clouds

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 5, May 2019

MOTIVATION

There are lots of user preference cloud Services for Data Storage but there is no Assurance for the secure data on cloud that why our system given some effective work to Cloud Server Provider for the purposed of Data Storing

II. REVIEW OF LITERATURE

1. “DEPSKY: Dependable and Secure Storage in a Cloud-of-Clouds”-Alysson Bessani, Miguel Correia, Bruno Quaresma, Fernando Andre and Paulo Sousa. Mining High Utility Pattern in One Phase without Generating Candidates. This article represents three variations of tree structure for high utility pattern mining for handling incremental databases. In this paper, three variations of tree structure have been proposed that are IHUP_L-tree, IHUP_{TF}-tree and IHUP_{TWU}-tree. These are very efficient for incremental and interactive high utility pattern mining. In that paper, the author used pattern growth approach to avoid the level wise candidate generation.
2. “NC-CLOUD: A NETWORK-CODING BASED DISRTIBUTED STORAGE SYSTEM IN A MULTI-CLOUD” – Dr. Mohammed Abdul Waheed, Sushmita B. N. Cloud is a proof of concept prototype of a network coding based storage system that aims at providing fault tolerance and reducing data repair cost when storing storage using multiple cloud storage. To provide fault tolerance for cloud storage, recent studies propose to stripe data across multiple cloud vendors. However, if a cloud suffers from a permanent failure and loses all its data, we need to repair the lost data with the help of the other surviving clouds to preserve data redundancy. We design a proxy based storage system for fault tolerant multiple cloud storage called NC-Cloud, which achieves reduction in repair traffic for a permanent multi-cloud failure. NC-Cloud is built on top of a network-coding-based storage scheme called regenerating codes. Specifically we propose an implementable design for the functional minimum-storage regenerating (FMSR) codes, which maintain the same fault tolerance and data redundancy as in traditional erasure codes (e.g., RAID-6), but use less repair traffic. We validate that FMSR codes provide significant monetary cost savings in repair over RAID-6 codes, having comparable response time performance in normal cloud storage operations such as upload/download.
3. “Scalia: An Adaptive Scheme for Efficient Multi-Cloud Storage”- Thanasis G. Papaioannou, Nicolas Bonvin and Karl Aberer . A growing amount of data is produced daily resulting in a growing demand for storage solutions. While cloud storage providers offer a virtually infinite storage capacity, data owners seek geographical and provider diversity in data placement, in order to avoid vendor lock-in and to increase availability and durability. Moreover, depending on the customer data access pattern, a certain cloud provider may be cheaper than another. In this paper, we introduce Scalia, a cloud storage brokerage solution that continuously adapts the placement of data based on its access pattern and subject to optimization objectives, such as storage costs. Scalia efficiently considers repositioning of only selected objects that may significantly lower the storage cost. By extensive simulation experiments, we prove the cost-effectiveness of Scalia against static placements and its proximity to the ideal data placement in various scenarios of data access patterns, of available cloud storage solutions and of failures.
4. “SPAN Store: Cost-effective Geo-Replicated Storage Spanning Multiple Cloud Services”-Zhe Wu, Michael Butkiewicz, Dorian Perkins, Ethan Katz-Bassett and Harsha V. Madhyastha. Cloud Services By offering storage services in several geographically distributed data centers, cloud computing platforms enable applications to offer low latency access to user data. However, application developers are left to deal with the complexities associated with choosing the storage services at which any object is replicated and maintaining consistency across these replicas. In this paper, we present SPAN Store, a key-value store that exports a unified view of storage services in geographically distributed data centers. To minimize an application provider’s cost, we combine three key principles. First, SPAN Store spans multiple cloud providers to increase the geographical density of data centers and to minimize cost by exploiting pricing discrepancies across providers. Second, by estimating application workload at the right granularity, SPAN Store judiciously trades

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 5, May 2019

off greater geo-distributed replication necessary to satisfy latency goals with the higher storage and data propagation costs this entails in order to satisfy fault tolerance and consistency requirements. Finally, SPAN Store minimizes the use of compute resources to implement tasks such as two-phase locking and data propagation, which are necessary to offer a global view of the storage services that it builds upon. Our evaluation of SPAN Store shows that it can lower costs by over 10 xs in several scenarios, in comparison with alternative solutions that either use a single storage provider or replicate every object to every data center from which it is accessed.

5. “Algorithms for Delta Compression and Remote File Synchronization”-Torsten Suel and Nasir Memon Delta compression and remote file synchronization techniques are concerned with efficient file transfer over a slow communication link in the case where the receiving party already has a similar file (or files). This problem arises naturally, e.g., when distributing updated versions of software over a network or synchronizing personal files between different accounts and devices. More generally, the problem is becoming increasingly common in many network based applications where files and content are widely replicated, frequently modified, and cut and reassembled in different contexts and packaging’s. In this chapter, we survey techniques, software tools, and applications for delta compression, remote file synchronization, and closely related problems. We first focus on delta compression, where the sender knows all the similar files that are held by the receiver. In the second part, we survey work on the related, but in many ways quite different, problem of remote file synchronization, where the sender does not have a copy of the files held by the receiver.

III. PROPOSED WORK

A. Proposed System Architecture

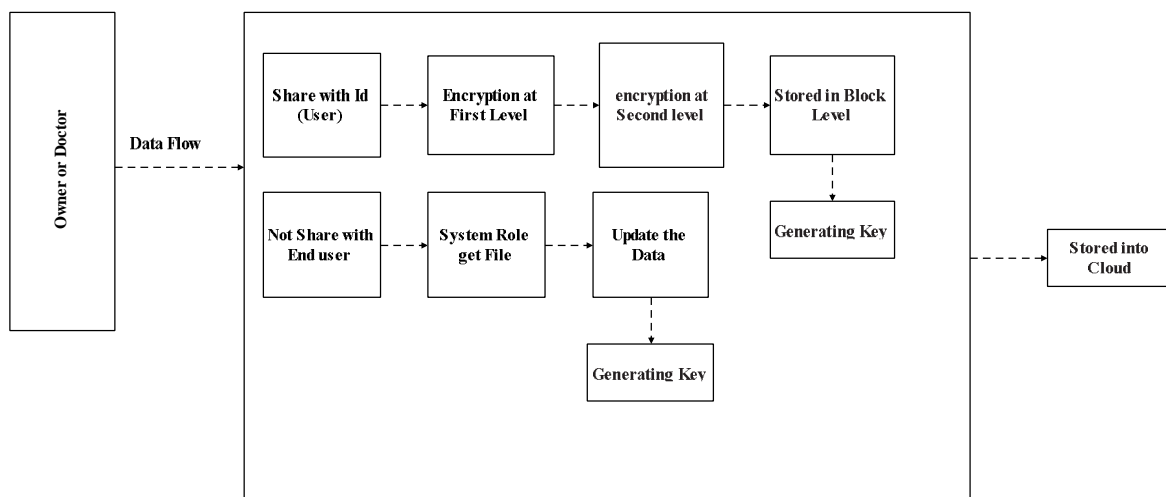


Fig.1: System architecture

B SYSTEM OVERVIEW

In Our Proposed System we used Stored Sim or storage planning and provide some security to user data, in that we use two layer protections such as LM Layer and CM layer all work is used in different layer. In our application de-duplication Layer and Chunking are performed in data in LM Layer and encryption and Bundling is performed in Data in CM Layer. Also we focus on access control, for the purpose for avoiding information leakage by using

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 5, May 2019

dynamic grouping concept's point of view also we give record monitoring to avoiding the attacker we use block chinning concept.

C ALGORITHMS

1. RSA Algorithms :- for content level encryption

INPUT: Required modulus bit length, kk .

OUTPUT: An RSA key pair $((N,e),d)$ where N is the modulus, the product of two primes $(N=pq)$ not exceeding kk bits in length; e is the public exponent, a number less than and coprime to $(p-1)(q-1)$; and d is the private exponent such that $ed \equiv 1 \pmod{(p-1)(q-1)}$.

- 1) Select a value of e from 3,5,17,257,65537
- 2) repeat
- 3) $p \leftarrow \text{genprime}(k/2)$
- 4) until $(p \bmod e) \neq 1$
- 5) repeat
- 6) $q \leftarrow \text{genprime}(k - k/2)$
- 7) until $(q \bmod e) \neq 1$
- 8) $N \leftarrow pq$
- 9) $L \leftarrow (p-1)(q-1)$
- 10) $d \leftarrow \text{modinv}(e, L)$
- 11) return (N,e,d)

2. Block Generating $(v, k+1, 1)$:- generating a new block :-

- 1) for $i=0; i \leq k; i++$ do
- 2) for $j=0; j \leq k; j++$ do
- 3) if $j=0$ then
- 4) $B_{i,j}=0$; else
- 5) $B_{i,j}=B_{i,j-1}+v$;
- 6) end if
- 7) end for
- 8) .end for
- 9) for $i=k+1; i \leq k+k; i++$ do
- 10) . for $j=0; j \leq k; j++$
- 11) . $B_{i,j}=0$ then
- 12) $B_{i,j}=[(i-1)/k + 1]$
- 13) Else
- 14) $B_{i,j}=B_{i,j-1} + \text{Mod } k+1(i-j+1)[(i-1)/k+1]$
- 15) . End if
- 16) End for
- 17) End for

3. Re-construction of B :- reconstruction of new Block :-

- 1) $E_0=B_0$; Steps 1
- 2) For $t=1; t \leq k+1; t++$ do
- 3) $E_t=B_{t-1}+E_{t-1}$ Steps 2
- 4) $B_t=[\text{Flag}]=1$;
- 5) $E_{t+1}=B[E_t, t/K]$ steps 2
- 6) $B_{t+1}[\text{flag}]=1$
- 7) End for
- 8) For $i=k+1; i \leq k+k+1; i++$ do

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 5, May 2019

- 9) If $Bi[Flag] \neq 1$ then
- 10) $Ebi[i+1/K] = Bi$ steps 3
- 11) End if
- 12) End For

IV. CONCLUSION

Distributing data on multiple clouds provides users with a certain degree of information leakage control in that no single cloud provider is privy to all the user's data. However, unplanned distribution of data chunks can lead to avoidable information leakage. We show that distributing data chunks in a round robin way can leak user's data as high as 80% of the total information with the increase in the number of data synchronization. To optimize the information leakage, Our design an approximate algorithm to efficiently generate similarity-preserving signatures for data chunks based on Min Hash using hashing Methods and also design a function to compute the information leakage based on these signatures. Next, we present an effective storage plan generation algorithm based on clustering for distributing data chunks with minimal information leakage across multiple clouds. New Methods are used Block Level whenever data are Clustering for distributing data chunks it will put into randomly blocks.

REFERENCES

- 1) A. Bessani, M. Correia, B. Quaresma, F. Andr e, and P. Sousa, "Depsky: dependable and secure storage in a cloud-of-clouds," ACM Transactions on Storage (TOS), vol. 9, no. 4, p. 12, 2013.
- 2) H. Chen, Y. Hu, P. Lee, and Y. Tang, "Nccloud: A network-coding-based storage system in a cloud-of-clouds," 2013.
- 3) T. G. Papaioannou, N. Bonvin, and K. Aberer, "Scalia: an adaptive scheme for efficient multi-cloud storage," in Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis. IEEE Computer Society Press, 2012, p. 20.
- 4) Z. Wu, M. Butkiewicz, D. Perkins, E. Katz-Bassett, and H. V. Madhyastha, "Spanstore: Cost-effective geo-replicated storage spanning multiple cloud services," in Proceedings of the Twenty-Fourth ACM Symposium on Operating Systems Principles. ACM, 2013, pp. 292-308.
- 5) T. Suel and N. Memon, "Algorithms for delta compression and remote file synchronization", 2002.