Reducing Cloud Storage Space Consumption using PRCR

L Raji, S.Thanga Ramya, A.Thilagavathy

Abstract: Data reliability constantly exists to be the important concern in dealing with the storage of data in distributed environment. For data reliability to be guaranteed, the replication approaches commonly used are in need of large amount of storage space which in turn leads to heavy cost for storage. The utmost need in data reliability requirement is the reduction of cloud storage space, In this paper, a comprehensive data reliability strategy is proposed so as to minimize the storage needed in the cloud and at the same time fulfilling the requirement of data reliability. The proposed model is based on Proactive Replica Checking for Reliability (PRCR) technique. The proposed method needs only a less amount of replication for ensuring data reliability which in turn leads to a more cost effective approach. Our proposed method also reduces the consumption of storage of data on the cloud

Keywords : cloud storage, data reliability, replication, PRCR.

I. INTRODUCTION

Cloud computing depends on partaking computing resources reasonably than needing native servers or individual devices towards the task of processing applications. Data reliability is a state that exist when data is sufficiently complete. In order to being reliable, data must also meet other tests for evidence. In this paper, we put forward a method that needs only a less amount of replication for ensuring data reliability which in turn leads to a more cost effective approach. Our proposed method also reduces the consumption of storage of data on the cloud.

The proactive kind of replication in data storage is vital in data reliability. For temporary data storage, a single replica for data reliability will suffice. For long-standing data storage, more than one replica is needed that is periodically checked for better reliability of data. The checking of the existence of the replication ensures their availability. The nature of the checking of the replication is proactive.

The major role of the proposed method exist in the development of a cost effective approach for ensuring data reliability that needs only a less amount of replication and thereby reducing the consumption of storage of data on the cloud. In case of any lost in replica, a recovery procedure is proposed. The recovery procedure generates new initialization of replicas.The proposed model exhibits the guarantee for data reliability even in the situation of failure among the storage requirements. The replicas needed are dynamically maintained. The proposed method uses less number of replicas thereby minimizing the storage of data in the cloud. In [1], reliability metrics was introduced to increase the storage system. The proposed system was able to recover quickly against failures occurred through hardware. In [2], fused backup technique was used to withstand faults that occur in data structures provided in the servers. The authors in [3] proposed a method to store and maintain very huge voluminous of data that are distributed in a very efficient manner. In [4], a reliability plan was proposed so that all the user's needs are guaranteed to be fulfilled by the cloud environment. The authors in [5] proposed the Carbonite replication algorithm intended for storage systems. In [6] , the factors related to reliability is dealt in detail and also gives the possible causes for the variation among the expected and the current reliability .In [7], the IT professionals inference in future is explained with lot of surveys that are to be predicted in near future. The authors in [8] proposed a cost efficient GRIDFTP server model for storage systems which can provide better reliability.

In [9], an approach to data integrity is proposed. Here small units of data are added together to form a big voluminous of data and integrating of data to its vendors are then performed. [10] dealt with problems associated with distributed hash table. Google file system was proposed in [11] that deals with very large voluminous of data and its processing towards solving real world applications. To guarantee data security in cloud, the authors in [12] proposed encryption and Obfuscation strategies. Confidentiality is maintained using these approaches. Moreover only authorized users are allowed to use the cloud data. Cloud storage management was discussed in [13]. Here the methodologies for efficient control of information into cloud and the storage of information into cloud were highlighted. Data storage capability using steganography and data obfuscation was proposed in [14]. The challenge involved around cloud computing and its future scope in the research possibilities were discussed in [15].

II. METHODOLOGY

A. Working of proposed system

The working of the proposed system is based on the following: File creation, replica management, scan process and recover file modules.

The advantages of the proposed are:
- ➢ The data reliability necessities and storage space time interval of diverse information identified.
- ➢ Involves intensive large scale data processing
- ➢ Both the computation and time overheads for encoding and decoding the data minimized.

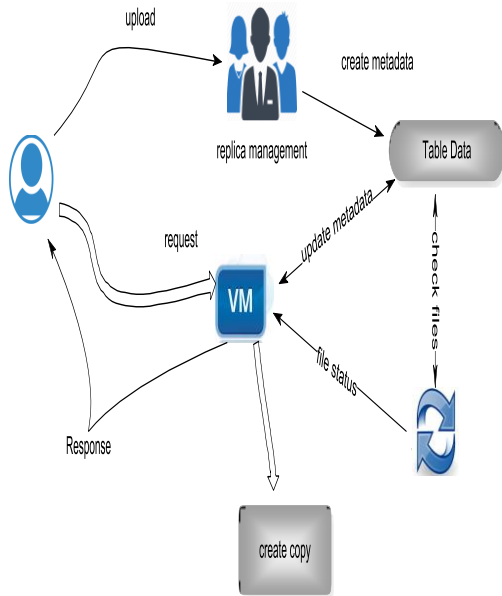Figure 1 illustrates the system architecture of the proposed system.

Figure 1: System Architecture

A. File Creation

User is given proper authentication to enter the system. Only authenticated customer are permitted for file creation. The files generated by the users are provided with a distinctive identifier. After the allocation of distinct identifier for the files, they are placed in the cloud storage. The files after the storage in the cloud will now be made accessible to the public users of the cloud. A replica is related through every file and is checked for its availability for better data reliability.
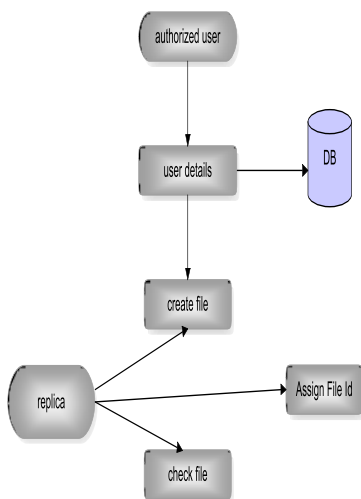

Figure 2: File Creation

B. Replication Management

After the generation of replicas for the file generated by the users, the numbers of replicas needed are determined based on the storage aspects of the file. Temporary files that are to be placed in the cloud storage need only one replica. Such replicas are generated and are checked by the model. Timestamp is associated with each replica generated for future checking. Long-term files that are to be placed in the cloud storage need more than one replica. The files are identified with the timestamp generated and their distinct file identifier.
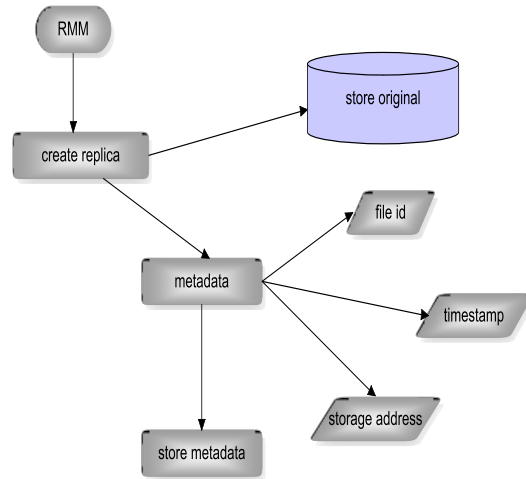

Figure 3: Replica Management

C. Scan Process

The scanning process of the files is done at certain time intervals. The same files when provided with two replicas are prone to the scan process. The files are checked whether they are temporary based or long-term based kind of storage in the cloud. The replica is identified with their address stored in the cloud. The scanning process ensures the file replicas are checked at certain time intervals and that the data reliability is maintained.
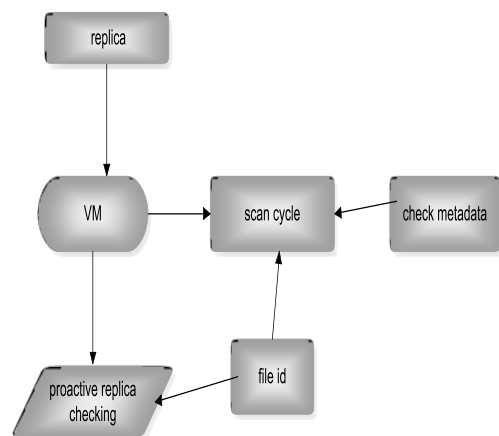

Figure 4: Scan process

D. Recover File

The virtual machine is provided with the replicas and performs the checking of the replicas. The timestamp and the period length of the time intervals are checked against the files identified by their distinct identifier. During proactive checking, if the replicas match, they are recovered. If the replicas do not match, they are replaced with new replicas. The timestamp and the time intervals are updated.
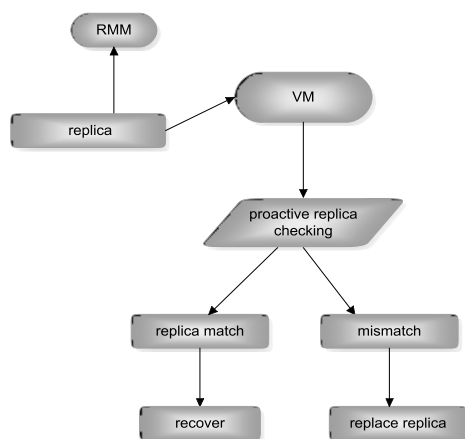

Figure 5: Recover file

III. CONCLUSION

In this paper, PRCR cost reliability management mechanism is presented. The PRCR strategy ensures the accessibility of each file for assuring that no files are there with more than two replica. Reliability management process using PRCR has shown that by using this mechanism, cloud storage consumption and cloud storage cost is reduced. The proposed method needs only a less amount of replication for ensuring data reliability which in turn leads to a more cost effective approach. Our proposed method also reduces the consumption of storage of data on the cloud.

REFERENCES

1. RekhaBachwani, LeszekGryz, Ricardo Bianchini, CezaryDubnicki, "Dynamically Quantifying and Improving the Reliability of Distributed Storage Systems", IEEE Symposium on Reliable Distributed Systems, pp. 85-94, 2008.
2. B. Balasubramanian and V. Garg, "Fault tolerance in distributed systems using fused data structures", IEEE Transactions on Parallel Distributed Systems, vol. 24, no. 4, pp. 701–715, Apr. 2013.
3. Byung-Gon Chun, Frank Dabek, Andreas Haeberlen, Emil Sit, Hakim Weatherspoon, M. FransKaashoek, John Kubiatowicz, "Efficient Replica Maintenance for Distributed Storage Systems and Robert Morris", 3rd Symposium on Networked Systems Design & Implementation, pp 45-58
4. E. Bauer and R. Adams, "Reliability and Availability of Cloud Computing". Piscataway, NJ, USA: IEEE Press, 2012.
5. B. G. Chun, F. Dabek, A. Haeberlen, E. Sit, H. Weatherspoon, M. F. Kaashoek, J. Kubiatowicz, and R. Morris, "Efficient replica maintenance for distributed storage systems," 3rd conference on Networked Systems Design and Implementation, 2006, pp. 45–58.
6. J. G. Elerath and S. Shah, "Server class disk drives: How reliable are they?" in Proc. Annual Symposium Reliability and Maintainability, 2004, pp. 151– 156.
7. J. Gantz and D. Reinsel, "Extracting value from chaos," IDC iview, vol. 1142, pp. 9–10, 2011.
8. A. Gharaibeh, S. Al-Kiswany, and M. Ripeanu, "ThriftStore: Finessing reliability trade-offs in replicated storage systems," IEEE Transactions on Parallel Distributed Systems., vol. 22, no. 6, pp. 910–923, Jun. 2011.
9. Eaton P, Weatherspoon H, and Kubiatowicz J," Efficiently binding data to owners in distributed content-addressable storage systems", 3rd International Security in Storage Workshop, Dec. 2005.
10. Freedman M J, Lakshminarayanan K, Rhea S, and Stoica I, " Non-transitive connectivity and DHTs", 2nd Workshop on Real Large Distributed Systems,(Dec. 2005.
11. Ghemawat S, Gobioff H, and Leung S T, " The Google file system" ,19th ACM Symposium on Operating System Principles ,Oct. 2003.
12. L. Arockiam, S. Monikandan, "Efficient Cloud Storage Confidentiality to Ensure Data Security", IEEE International Conference on Computer Communication and Informatics (ICCCI), 2014.
13. Zhan Ying, Sun Yong, "Cloud storage management technology", Second International Conference on Information and Computing Science, 2009
14. B. Fathima Mary ; D.I. George Amalarethinam, "Data Security Enhancement in Public Cloud Storage Using Data Obfuscation and Steganography", World Congress on Computing and Communication Technologies ,2017.
15. Khushboo Jain , Vinod Maan, " A Review Paper on Considerations and Challenges in Cloud Computing " International Journal of Computer Science and Mobile Computing, vol.6 Issue.4, pp. 363-368, 2017.

AUTHORS PROFILE

L.Raji received her B.E. (Computer Science and Engineering) from University of Madras , M.E(Computer Science and Engineering) from Anna University and pursuing Ph.D in Anna University. She is working as Assistant Professor of Computer Science and Engineering at R.M.K. Engineering College. Area of interest is cloud computing.

Dr.S.Thanga Ramya, B.E, M.S(by Res), Ph.D, is an Associate Professor in the Department of Information Technology, since June 2008. She obtained her B.E(CSE) from Dr.Sivanthi Aditanar College of Engineering and M.S by Research (ICE) from Anna University, Chennai. She has obtained her Ph.D in Information and Communication Engineering from Anna University, Chennai, in 2017. She has been in the teaching profession for the past 15 years and has handled both UG and PG programs. Her areas of interest include programming languages, database management and data mining. She has published 7 papers in various International Journals and Conferences. She has attended many workshops & FDPs sponsored by AICTE related to her area of interest. She has also published a book - "Fundamentals of Computing" for the benefit of first year B.E/B.Tech students of Anna University. She is the life member of ISTE.

A. Thilagavathy received her B.E. (Computer Science and Engineering) from University of Madras, M.E. (Computer Science and Engineering) from Anna University and pursuing Ph.D in Anna University. She is working as Associate Professor of Computer Science and Engineering at R.M.K. Engineering College. Area of interest is Image Processing and soft computing.