International Journal of Advanced Research in Engineering and Technology (IJARET)
Volume 12, Issue 1, January 2021, pp. 354-361, Article ID: IJARET_12_01_032
Available online at http://www.iaeme.com/IJARET/issues.asp?JType=IJARET&VType=12&IType=1
Journal Impact Factor (2020): 10.9475 (Calculated by GISI) www.jifactor.com
ISSN Print: 0976-6480 and ISSN Online: 0976-6499
DOI: 10.34218/IJARET.12.1.2021.032

A SURVEY OF SECURE CLOUD STORAGE MECHANISM: CHALLENGES AND ITS SOLUTIONS

Banoth Anantharam

Research Scholar, Department of Computer Science and Engineering, Sri Satya Sai University of Technology & Medical Sciences, Sehore, Madhya Pradesh, India.

Assistant Professor, Department of Computer Science and Engineering, Keshav Memorial Institute of Technology, Narayanaguda, Telangana, India.

ABSTRACT

Cloud storage has been a boon for many organizations and the individuals as it reduces the burden of safety and security of their data, in addition to minimizing the investment for infrastructure. Every cloud service provider offers not only the storage service but it also extends to preserve the data privacy and security. The inherent mechanism that many cloud service providers implement is n-backup by which the client is guaranteed for data recovery in the case when the original copy of data is damaged or lost. However, for each of the backup copy, it is evident that more storage space is required. In this digital big data era, industry is experiencing the data explosion hence more space and infrastructure would become a necessary need for the service providers. In this paper, a study is conducted on various cloud storage mechanisms, challenges and identified the gaps. Finally, possible solutions are mentioned which is a need for the present cloud storage scenario.

Key words: cloud computing, cloud storage, data centre, storage mechanisms, storage backup.

Cite this Article: Banoth Anantharam, A Survey of Secure Cloud Storage Mechanism: Challenges and its Solutions, *International Journal of Advanced Research in Engineering and Technology*, 12(1), 2021, pp. 354-361.
http://www.iaeme.com/IJARET/issues.asp?JType=IJARET&VType=12&IType=1

1. INTRODUCTION

The fundamental concept of Cloud Computing to perform the user tasks in distributed, parallel mode by using appropriate technologies through which the tasks can be completed at a faster pace. In the cloud computing paradigms, Infrastructure-as-a-Service (IaaS) plays a major role with the storage component. To offer more robust storage mechanisms, cloud storage providers offer n-backup mechanism. The mechanism involves number of backup copies for an original

copy of data range from two to three. The main objective of offering such n-backup mechanism is to create the fault-tolerance and resilience for the applications as well as the owners of the data. Apparently, the cloud storage can be used for various purposes such as backup and recovery, software test and development, cloud data migration, building big data and data lakes, etc. Moreover, cloud storage requirements are availability, data integrity and security. Considering the artefacts of the cloud storage, Figure. 1,

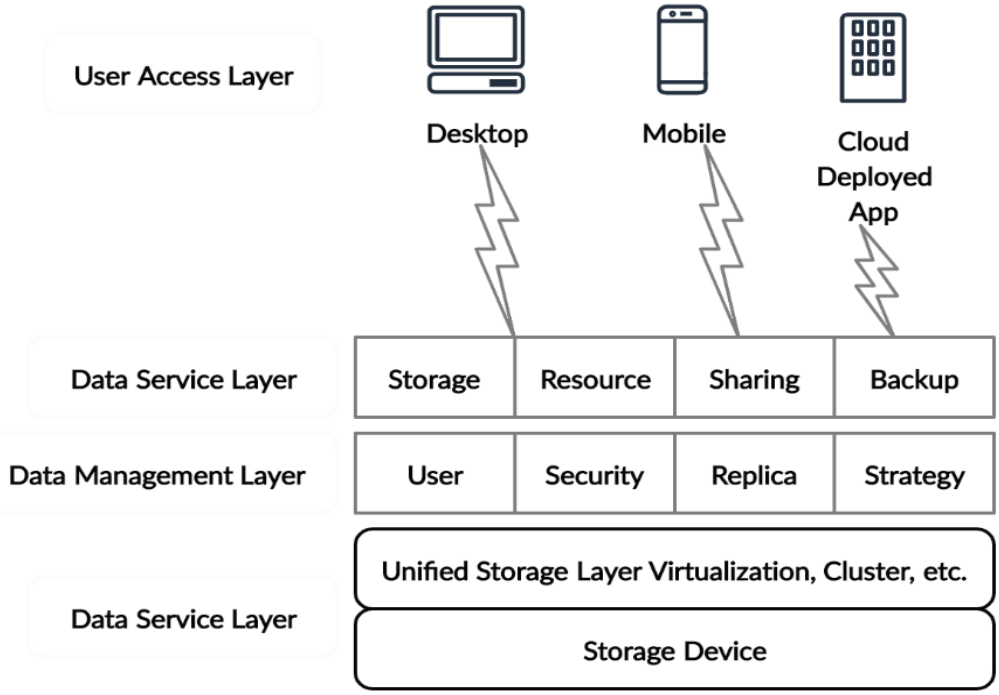

Figure 1. Architecture of Cloud Storage

the user would get benefitted to the maximum extent in the context of the reducing the resources needed to store the data, deploy the system and maintain the data.

2. PROBLEM STATEMENT

The work proposed in this paper if for an optimized and secured storage solution, for the cloud storage service providers, which shall reduce the burden on the storage systems. From the Figure 2 it can be observed that while storing the data in the cloud, the system experience following issues:

2.1. Load on Storage System

In case of the small-sized objects maintaining extra copies doesn't create much storage load on the cloud storage servers, but if the objects are larger in size then the data redundancy creates an exponential growth in the requirement for cloud storage resources which not only increases the cost of maintenance but also consumes lot of physical space required to implement the cloud storage servers. Though the cloud service providers have projected the reason of cooling the system but dearth of space on earth is also a potential reason to setting up the cloud system in the ocean such Microsoft did it in the Pacific Ocean.

2.2. Load on Networking Resources

Two types of network resources are required while the data has to be stored in the cloud storage: data generated by the external system, and the data generated in the cloud application. If the data is generated in the external system to be stored in the cloud storage then the operation

requires two kinds of networking resources: user network-to-cloud network, cloud network-to-cloud network. If the data is generated in the cloud application then user-level networking resource is not needed whilst internal cloud network itself is sufficient. However, in both the scenarios, much of the load that is generated in the internal cloud network would be tripled for every chunk of the data to be stored.

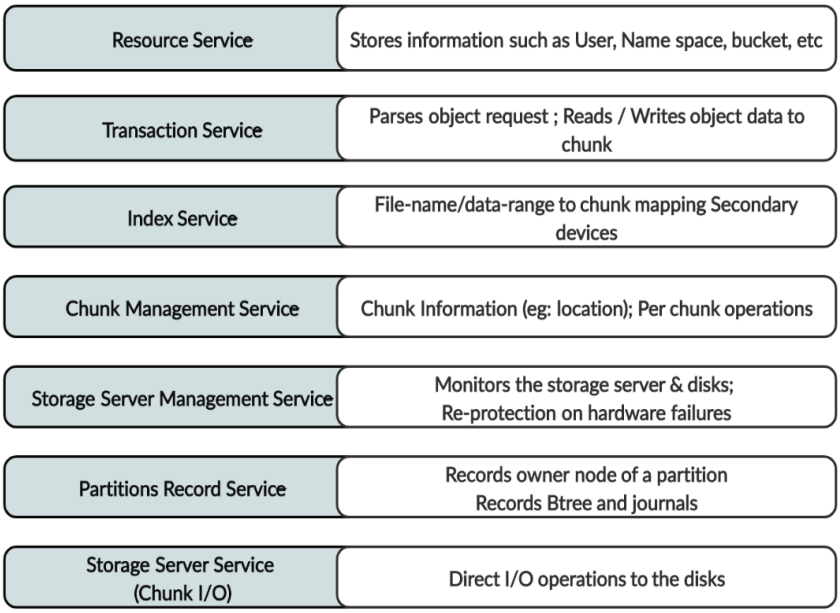

Figure 2. Storage Engine Services adopted by Dell-EMC for Elastic Compute Service (ECS)

2.3. Load on Storage-Connected Computational Resources

For every object to be stored in the cloud it needs to compute the amount of storage required for the object vs the availability of storage devices. At the same time, the computation time is also tripled for every object to find the three storage units, one for the original file and two others for backup in the context of future recovery which is the part of the definition of the cloud.

3. LITERATURE REVIEW

Generally, large-scale data centres implement remote backups for the improvement of system reliability for achieving the efficient fault tolerance. But considering the distance of the remote locations which result into transmission cost, Yu Hua et.al. [1] have et.al. have proposed a cost-efficient method for remote backup services in the context of Enterprise Clouds, called Neptune. As part of the work, the authors have developed a filtration scheme to transmit the data to the long-distance data centers by eliminating the redundancy. The reduction of redundancy is achieved by incorporating the method of compressing the similarity of files. It further incorporates deduplication at chunk-level and uses delta-compression for compressing the files. In order to reduce the overheads and complexity, Neptune makes use of locality-aware hashing for grouping the similar files and suggests shortcut delta chains for fast remote recovery.

Before developing any solution for the cloud storage technology, it is inevitable to understand the architectures that are implemented successfully for in the existing scenario. Yuhuan Qiu et.al. [2] have presented concepts and architecture models that are implemented for the cloud storage in most of the cloud solutions. In the process the authors have presented the features and advantages of which space leasing, reducing the management difficulty, data replication and backup, scalability are elevated to be the most prominent and important. The

types of cloud systems vary with the need and the customer orientation such as individual-level storage requirement to industry-level storage requirement. Hence, depending upon on the scale of requirements, and the context for which the storage solution is designed. iCloud, Baidu cloud disk have been discussed along with the issues like broadband bottlenecks, infrastructure, security and availability.

When the services are offered by the cloud service providers include Infrastructure-as-a-service then they have to concentrate more on number of components for IaaS's functionality. Additionally, the functionality has to be in accordance with huge system duplication of components but it increases the cost of maintainability and operability. S.R.Ali [3] have analysed the cloud computing reliability in the context of unified infrastructure and proposed the methods to implement it a reduced cost. The solution should also provide the operator that consists of computing elasticity which helps the functionalities to grow or de-grow the cloud based on services on-demand. Because the clouds are located at various geographical locations, Single Point-of-Failure (SPOF) can be considerably reduced. As part of the analysis server virtualization has been studied in the context of hypervisors, various virtual states, transition states for which Virtual Machine Markov model has been studied. Apart from them, VM recovery mechanisms, cloud failover, container virtualization, and reliability analysis of VoIP in cloud environment have been discussed.

Zhang et.al. [4] have designed and implemented data backup and recovery system with security enhancements which mainly focuses on mainly focus on the availability including the confidentiality in the backup operation as well as recovery operation. The work has considered various modules in the architecture of the system such as backup module, recovery module, network communication module, logging and transmission module, task management module which exist at the upper layers of the architecture. The key module i.e., security enhancement module, consists of various modules such as identity authentication, operation approval, account system, log audit, data protection, backup data management, and policy management. The system communication has been defined for the communication between client, web console and the backup and recovery server. Through their work the authors suggest that when user authentication is strengthened, verification of the legitimacy between the data and target computer, the process control during business operations and the audit of the key operation, the system can prevent administrators from backing up data to illegal disk and recovering backed up data to illegal target machine and other similar risk of data leakage.

To achieve the budget association and savings, efficiency of government agencies' business-processes, one of the mechanisms is to implement a new information model which is migration-based for the benefit of cloud computing, ICT-outsourcing and consolidating the orders. The key requirements were proposed by Aubakirov et.al. [5] to support the open standards for storage and distribution of virtual machines; Logical partitioning of resources pools provided by virtualization platform on virtual computing data centers with fixed service quality; solutions on provision of network safety, etc. In the process the authors have defined various subsystems such as services, resources virtualization, computing platform, data transfer, data storage, communication, data centers, management and monitoring, information safety, backup and recovery, and technical assistance. The authors have observed that the computing capacities' efficiency per kilowatt-hour will increase, which in-turn will lead to the increase in environmental friendliness of government agencies performance.

In the work submitted by Odun-Ayo et.al. [6] have presented various aspects involved with data storage in cloud computing. The authors have perceived the key issues that are part of cloud storage services are deployment, virtualization and availability, data organization, data migration and load balancing, data deduplication, and security. Furthermore, they have listed the security concerns such as data privacy and integrity, data recovery and vulnerability,

improper media sanitization, data backup, and data outage. In the process of exploring the key features the work has included the comparison of IBM cloud object storage vs Amazon S3 for which single/multi-tenancy options, deployment options, customization and control, and API support were compared. The results of the comparison has resulted to diversity in which IBM cloud object storage has produced better throughput than Amazon S3, while read and write latencies were much lower than S3.

Outsourcing of matrix multiplications tasks at a larger scale to many distributed servers or cloud is necessary to increase the speed of computation. But security is a serious concern when these servers are not trustworthy. Chang et.al. [7] have studied the problem of secured and distributed version of matrix multiplication from the view point of servers which are not trustworthy over distributed servers. This issue is categorized to be part of secure function computation and has established substantial attention in the cryptography community. But, portraying the essential bounds of information-theoretically secure matrix multiplication still remains as an open issue. Their focus was on information-theoretically secure distributed matrix multiplication with an objective of portraying the minimum communication overhead. The ability of secure matrix multiplication was defined as "the maximum possible ratio of the desired information and the total communication received from N distributed servers". After the investigation of single-sided and fully secure matrix multiplication issues the authors have proposed a scheme based on secret sharing for the single-sided secure matrix multiplication model. Furthermore, two interesting open problems have been proposed: find a converse (upper bound), for a fully secure matrix multiplication issue; and generalizing these ideas for other secure distributed computation tasks.

Security and privacy are the most critical issues that need to be addressed in designing a computing environment that is reliable and trustworthy. Tariqul et.al. [8] have classified and characterized the various security and privacy challenges associated with cloud computing. In the process they have conducted a survey on three well-known cloud security frameworks namely ENISA, CSA, and NIST that aim to provide a compilation of risks, vulnerabilities and also the best practices to resolve them. These three entities provide a comprehensive overview of the current security, privacy, and trust issues, and thus, help in understanding the current status of Cloud security. They have also presented a variety of security and privacy concerns associated with Cloud Computing, identified major threats and vulnerabilities, and classified them into six categories: Network Security, Virtualization and Hypervisor Security, Identity and Access Management, Data and Storage Security, Governance, and Legal and Compliance issues. Each of these categories identified several threats and vulnerabilities, resulting in further classification. The authors have observed that it is evident that for the wide spread adoption of the cloud, these issues must be addressed thoroughly. Therefore, enrichment of the existing solution techniques as well as more innovative approaches need to mitigate these problems are needed.

The latest technological developments have ignited the popularity and success of the cloud. Since it provides cost-efficient architectures that provide the transmission, storage, and exhaustive computing of data, this new paradigm is gaining an increasing interest. Kaaniche et.al. [9] have conducted an exhaustive study to deliver a consistent assessment about both data security concerns and privacy issues that are faced by clients in cloud storage environments. This work brings a critical relative analysis of cryptographic defense methods. Apart from this, their work has explored research directions in addition to the technology trends to report the protection of outsourced data in cloud infrastructures. The foundation for the their work is that the storage services carry a good number of challenging design issues, significantly because of both loss of abstract nature of clouds and data control. The authors are in a perception that there exists no cryptographic mechanism that can be installed to guarantee privacy of users, for the

settings in data sharing. In the process it was also found that Fully Homomorphic encryption was a proven method as unable to implement privacy demands of the common cloud services. It is also observed that the schemes of FHE cannot be completely private due to their inability to support circuit complication. Therefore, it has been deduced that computations over muti-client inputs needs reliable hardware. Following this approach, it is assured that Secure Multiparty Computation (SMC) can be capable resolution to privacy-preserving processing. In general, SMC mechanisms permit quite a lot of clients to collaboratively exicute any function, over private inputs. This makes the distribution of trust to be possible across many entities.

One of the potential issues in privacy and security is stimulated by the fact of likelihood that cloud operators can access the user's sensitive data. This concern intensely rises users' apprehension and decreases the adoptability of cloud computing in many of the industries, such as the governmental agencies and financial industry. The work proposed by Yibin Li et.al. [10] has focused on this issue and proposed an approach namely, intelligent cryptography by which the cloud service operators would not be able to directly reach partial data. Their approach splits the file and stores the data separately in the cloud servers which are distributed across various locations. They have designed an alternative approach for determining whether the data packets require a split to condense the operation time. They have named the proposed scheme as Security-Aware Efficient Distributed Storage (SA-EDS) model, which are supported by their algorithms. These algorithms include: Efficient Data Conflation (EDCon) Algorithm, Secure Efficient Data Distributions (SED2) Algorithm and Alternative Data Distribution (AD2)Algorithm. The experimental calculations have evaluated both security and efficiency performances. Moreover, the investigation results represent that their approach can successfully protect main threats from clouds and needs with an acceptable computation time.

4. RESEARCH GAP

Although there's an extensive research that was conducted to reduce the load on the storage systems, there's a lot of scope to develop more effective solutions that can address to optimize the storage usage. However, most of them are hypothetical which are not practically possible to be implemented in the cloud scenario. A systematic study in the context of storage optimization would reveal good number of gaps that which are potential to enhance the solutions. A solution as a Proof-of-work, in terms of optimization which does not degrade the overall performance, if not enhanced, would always be adoptable by the cloud service providers. In the context of cloud storage most of the researches have proposed the techniques to compress the huge volume of data, and data organization methods for the data that was generated by the large-scale users. Whether the lossy or lossless compression, every compression technique would stop after certain stage when it reaches the trade-off, i.e., the computational resources required to compress the huge-volume of data vs the time required to compress. These techniques not only degrade the overall performance but also consumes lot of energy and increase the computational costs which in turn would increase the cost for the user. Cloud backup systems have sophisticated requirements on reliability than common backup software. However, users unconsciously feel it safer to backup critical data on visible devices. Doubts may arise about the security of backing up private data in faraway data centers.

Cloud backup poses various abnormalities, as most of the providers concentrate on providing security for the cloud backup than the backup software by which the security issues would be as a result. Hence, the security for the data backup is also an essential activity.

5. RESEARCH CHALLENGES

While implementing the solutions which were suggested by the researchers the industry and the research teams have experienced many issues. These issues range from the data size to the

technologies adopted while implementing the solutions at the data centre. Though many could be solved, but the challenges as mentioned below needs a special focus.

- Before the data is backed up, more time has to be spent on identifying the requirements of the file system at the storage.
- While the data is being backed up, Though the data security is vital, backup service security is at the stake which would cause the fatal damage to the data.
- An elastic security service is needed for the cloud service provider because the types of cloud data storage: object storage, file storage, and block storage.
- Though high-performance infrastructure is available with the cloud storage service providers most of the resources are consumed for the scalability and elasticity which needs to invoke various functions to be executed at the time of redundant backups.
- An asynchronous procedure is adopted for Cross-region replication to backup each object in the storage, this would cause a requirement of computation of location to be identified and then the creation of meta data information.
- Cloud service provider provides the data security in two modes, client-side encryption and server-side encryption. Client-side encryption ensures that the data is encrypted before it reaches the cloud server, whereas the Server-side encryption ensures that the data is encrypted without the client having the encryption facility. However, except the large-scale organizations most of the small-scale organizations and individuals would not prefer the client-side encryption. At the same time, there are apprehensions across the users about the integrity of the security services that are provided at the server-side.
- Because of the distributed architecture of the cloud servers elastic file system facility, higher throughput can be experience if the I/O workloads are larger. For the lower I/O workloads the throughput is not constant and is guaranteed to be higher.
- If the application at the client-side is not parallelizable across multiple instances, then the throughput cannot be at a better number as the data transfer is serialized.
- Write performance and consistency of data transfer speed is also dependent on whether the client system is able to support the asynchronous writes to the file system.
- Identity and Access Management permission for API calls to access the backup software, security groups for the cloud instances and file system permission for user-level, and group-level must be defined appropriately.

6. CONCLUSION

A study on various cloud storage mechanisms was conducted and presented the observations, in this paper. Most of the implementations have adopted the n-backup mechanism to ensure the recovery in the case of any damage or data lost in the original copy. It is observed that a potentially extra storage space, infrastructure and related business logic have become a necessary need to implement n-backup mechanism. As part of the study few issues have been identified which ought to be addressed in order to reduce the storage space burden on the service providers. Apparently, research challenges have also been identified while implementing the solutions by the service providers and presented the same in the paper.

7. SUGGESTIONS

Developing the solutions for a problem that potentially degrades the performance of any system needs an immediate attention. From the discussion in this paper it is evident that consumption of more space and investment for infrastructure would always be a challenging issue for the cloud storage service provider. Hence, there must be the development of solutions which reduce

the burden of extra storage space without compromising for the privacy and security. These solutions shall not only reduce the storage load but also improve the communication efficiency between the client and service provider.

REFERENCES

[1] Hua, Yu, Xue Liu, and Dan Feng. "Cost-Efficient Remote Backup Services for Enterprise Clouds." *IEEE Transactions on Industrial Informatics*12, no. 5 (2016): 1650-657. doi:10.1109/tii.2016.2543258.

[2] Yuhuan, Qiu. "Cloud Storage Technology." *Big Data and Cloud Innovation*1, no. 1 (2017). doi:10.18063/bdci.v1i1.508.

[3] Ali S.R. (2019) Cloud Computing Reliability Analysis. In: Next Generation and Advanced Network Reliability Analysis. Signals and Communication Technology. Springer, Cham

[4] Zhang, Jianping, and Hongmin Li. "Research and Implementation of a Data Backup and Recovery System for Important Business Areas." *2017 9th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*, 2017. doi:10.1109/ihmsc.2017.209.

[5] Aubakirov, Margulan, and Evgeny Nikulchev. "Development of System Architecture for E-Government Cloud Platforms." *International Journal of Advanced Computer Science and Applications*7, no. 2 (2016). doi:10.14569/ijacsa.2016.070235.

[6] Odun-Ayo, Isaac, Olasupo Ajayi, Boladele Akanle, and Ravin Ahuja. "An Overview of Data Storage in Cloud Computing." *2017 International Conference on Next Generation Computing and Information Systems (ICNGCIS)*, 2017. doi:10.1109/icngcis.2017.9.

[7] Chang, Wei-Ting, and Ravi Tandon. "On the Capacity of Secure Distributed Matrix Multiplication." *2018 IEEE Global Communications Conference (GLOBECOM)*, 2018. https://doi.org/10.1109/glocom.2018.8647313.

[8] Islam, Tariqul & Manivannan, D. & Zeadally, Sherali. (2016). A Classification and Characterization of Security Threats in Cloud Computing. International Journal of Next-Generation Computing. 7.

[9] Kaaniche, Nesrine, and Maryline Laurent. "Data Security and Privacy Preservation in Cloud Storage Environments Based on Cryptographic Mechanisms." *Computer Communications* 111 (2017): 120–41. https://doi.org/10.1016/j.comcom.2017.07.006.

[10] Li, Yibin, Keke Gai, Longfei Qiu, Meikang Qiu, and Hui Zhao. "Intelligent Cryptography Approach for Secure Distributed Big Data Storage in Cloud Computing." *Information Sciences* 387 (2017): 103–15. https://doi.org/10.1016/j.ins.2016.09.005.