International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 06 | June 2019 www.irjet.net p-ISSN: 2395-0072

An Integrity Auditing &Data Dedupe with Effective Bandwidth in Cloud Storage

Smita Anandrao Patil [1], Gayatri Londhe [2]

[1]*Assistant Professor, Dept. of E&TC Engineering, Dr. D.Y. Patil Institute of Technology, Pune, Maharashtra, India*
[2]*Assistant Professor, Dept. of E&TC Engineering, Dr. D.Y. Patil Institute of Technology, Pune, Maharashtra, India*

Abstract - *Cloud computing is said to be the biggest thing from the beginning of the era of internet. In fact, cloud computing could be the greatest technical breakthrough anyone can experience in the lifetime. However, adopting cloud computing for any business is a complex decision that involves many aspects. The key factors are security of data, privacy, compliance to regulatory requirements, performance in a multi-tenant environment and 24/7 availability. Cloud computing makes use of applications, development platforms and system hardware in remote datacenters to deliver services over the internet. Cloud computing tightens the security by regulating user authentications, by using cloud resources and by monitoring user trace and activities. Cloud computing allows its users to access, use and store the data. To manage this data, to check the integrity of this data and to provide security to this data a secure system is needed. Third party auditor (TPA) checks the integrity of data. A well-known hashing algorithm, standard hashing algorithm is used for checking the existence of data which is helpful for checking the deduplication. Deduplication is a data compression technique which discards the duplicate copies of data having same content. Also Advance Encryption standard algorithm is used for enhancing the privacy and security of data.*

Key Words: *TPA; SHA-512; MHT; BANDWIDTH*

1. INTRODUCTION

Cloud is a source where data owners have a convenient, reliable, on-demand access and it combines the resources, such as servers, storage and applications over the internet. Users don't have a control of underlying hardware infrastructure that is owned and managed by the provider. They access the services or allocated resources by using a web browser. Cloud computing provides suitable, on request network which can be disclosed with essential Management effort or gives best service interaction. It can be provisioned promptly to combined resources such as storage, applications, services, networks and servers.

Although cloud computing can accumulate very large data, it fails to provide network bandwidth and save disk space because two or more clients upload the same file. To overcome this problem, a well-known data compression technique termed as deduplication is used. In data deduplication, rather than storing number of data copies having same content, server removes duplicate data copies by storing only one physical copy and referring other data copies to the already stored data copy. The Third party auditor is a kind of inspector. In cloud environment anyone can access the data over the internet. Hence user authentication and access control is very important in the cloud. Trusted third parties are independent service providers and are assumed to have a certain level of trust.. It is very important to provide public auditing service for cloud data storage, so that the user trusts an independent third party. The Trusted Third Party (TPA) checks the integrity of data by taking the current (this value is sent by service provider to TPA which is a live hash value) and actual hashing value (this value is sent by the user) of the file and gives response to user

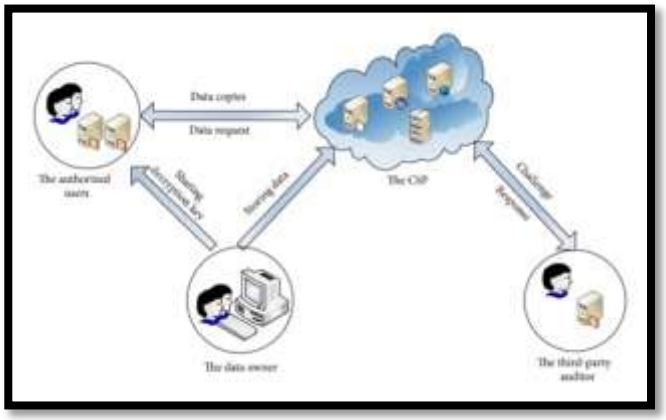

Fig -1: Cloud Storage System

2. BACKGROUND THEORY

User/Client- It is one of the element of cloud who is going to store his/her data or files on cloud space. he/she can access or modify data at any time from any place and anytime.

CSP/Admin-It is the owner of the cloud who is going to provide storage space and on demand services to the users who is registered with the cloud. He is the one who is responsible for issues related to user data.

TPA-Third Party Auditor- This is the main element of cloud who works on behalf of the client.TPA is checking integrity of data and notify the status of data when user send a request for integrity checking of Files or data. It reduces the burden of user by taking care of data

3. PROPOSED SYSTEM

A model is prepared in which client, a cloud service provider (CSP)/cloud server and TPA are present. Cloud user is the one who stores large amount of data or less on a cloud server. Users are the devices that the end users interact with to manage their information on the cloud. Cloud server is a place where we are storing cloud data and that data will be managed by the cloud service provider. Third party auditor acts as an inspector which varies the users authentication and also checks the integrity of data. Performance and efficiency is better than existing systems as we are using two TPAs. i.e. if main TPA goes offline secondary TPA will take care of user data. As user cannot store data directly on cloud server thus It is encrypted with AES-128 algorithm which is faster in encryption and decryption than RSA and it requires less storage requirement. SHA-512 and MHT (Merkle Hash Tree) algorithm are used for authentication and data integrity checking. With this efficient data integrity checking it also provides data recovery in case when data is corrupted. Also user can edit his/her file any time and all the updating made by user stored on clouds so that after modification done by user if he want his original file that file can also recover with this scheme. File is divided into different fragment and each fragment is secured with key so that it is difficult to hack file. With this proposed scheme also supports auditing for dynamic data as well as support dynamic operation on data. Proposed system model is divided into different phases

1] Uploading/downloading of file

2] Authentication of file

3] Integrity checking of file

4] De-dupe of data

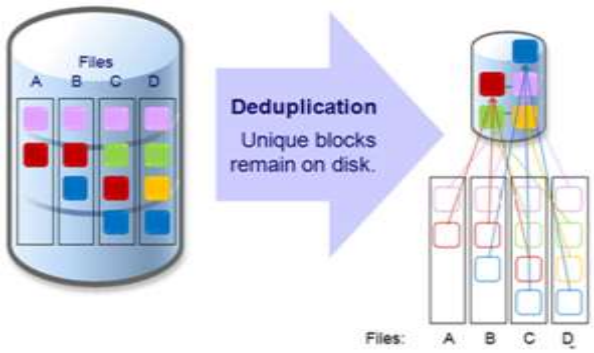

Fig -2: Data De-duplication

Byte level de-duplication:

Byte-level deduplication, as shown in Figure 2.2, doesn't need extra processing. In this scenario data groups are analyzed in the earliest way, byte by byte. For unnecessary groups it works more efficiently. Time required to perform Byte-level deduplication is more as compared to other levels.

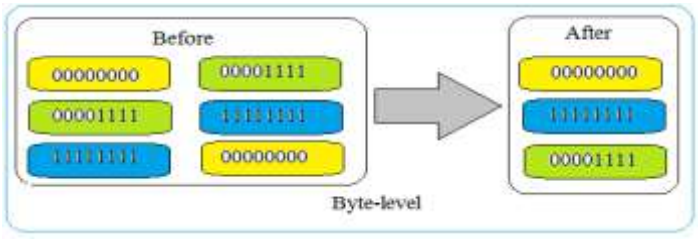

Fig -3: Byte level de-duplication

Block level de-duplication:

The problem mentioned above can be resolved by using a block level data de-duplication method as shown in Figure since, it does not have such a drawback. At the block level data de-duplication, many groups of block having fixed or variable length are formed by splitting a file. If any minor change is made in large file then rather than storing all fragments, the system stores its changed fragment only.

Block level de-duplication requires more processing power as compared to file level de-duplication since there is a great increase in the number of hash values that need to be processed. In the same way, index of block level for tracking individual block increases in large scale. Complexity is more in case of variable length blocks. It may possible that for two different data fragments the same hash number may be generated, which is called as a hash collision. The system will not save any new data if such thing happens because it observes that, index already have that hash number. According to environment requirements variable block size can be specified 4 or 256 Kb or any other. In this, block sizes having larger size requires fewer comparisons, but in some cases it provides minimum compression and more comparisons are required for smaller one as it provides better compression.

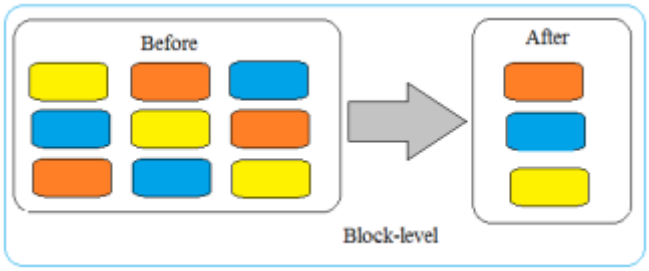

Fig -4: Block level de-duplication

File level de-duplication:

File level de-duplication, can be performed without difficulty as shown in figure Entire files hash number is easy to generate by using hashing algorithm. Hence, it requires fewer computations. However, limitation of this method is, Hash number will change even if there is a change in single byte. This forces cloud server to save both file versions into cloud storage.

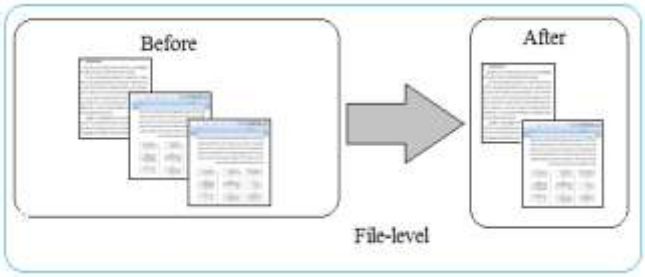

Fig -5: File level de-duplication

4. PERFORMANCE ANALYSIS

Figure shows comparative analysis of storage space in existing system and implemented system. On x-axis no. of files are shown and on y-axis size in mb is shown. In first case 10 files are considered and in second case 20 files are considered. Existing system does not remove duplicated files as such it requires more space as compared to implemented system. In implemented system duplicated copies are stored only once. Table shows data of storage space.

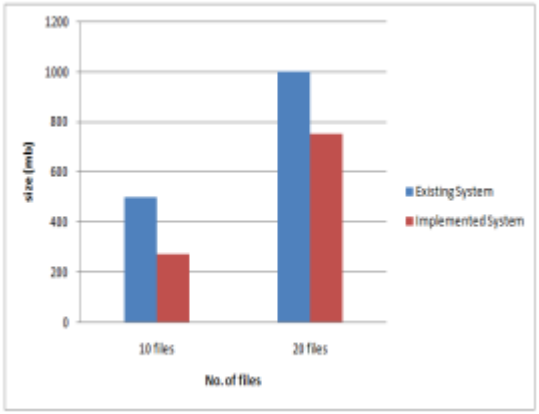

Chart -1: comparative analysis of storage space

Figure shows time required for auditing by CSP and TPA. We took files of size 130 mb, 190 mb and 460 mb. In all three cases, time required to audit file by TPA is more. Table shows data of file auditing.

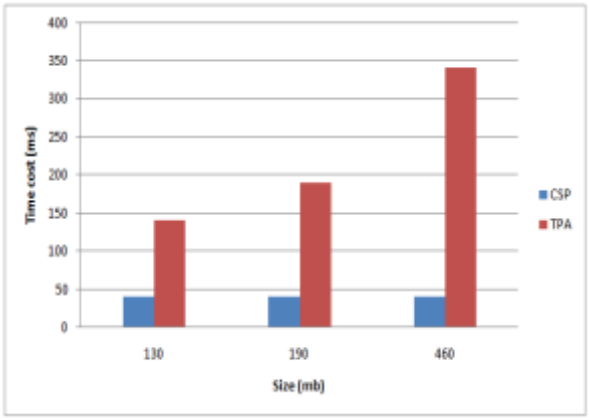

Chart -2: Comparative analysis of Time Requirement

Sample paragraph Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Abbreviations such as IEEE, SI, MKS, CGS, sc, dc, and rms do not have to be defined. Do not use abbreviations in the title or heads unless they are unavoidable.

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper.

5. CONCLUSION

A system is implemented for security of user's data when data is stored into the cloud. Preserving privacy in a cloud environment is done by using convergent encryption method. To protect the data from unauthorized access and to ensure that data is intact, the TPA model implemented a scheme, which solves a problem of integrity, unauthorized access, privacy and consistency. In implemented system there is intervention of TPA which reduces the overhead of user. While uploading the file in cloud, TPA generates the hash value for checking the deduplication of that file. TPA does not leak any side channel information. In implemented system, a system is designed having intervention of TPA and using SHA for hash value and AES for encryption. Aiming to achieve securing cloud files and deduplication, a better encryption method AES is being used. SHA generates strong message digest which is more secure. Also because of using convergent encryption method, system has become more secure.

REFERENCES

[1] Jingwei Li, Jin Li, Dongqing Xie and Zhang Cai, "Secure Auditing and Deduplicating Data in Cloud", in Proceedings of the IEEE Transactions on Computers, Vol. 65, No.8, Pp. 2386 - 2396, 2016.

[2] M. Armbrust, A. Fox, R. Griffth, A. D. Joseph, R. Katz, A. Konwinski, G. Lee,D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, "A view of cloud computing,"Communication of the ACM, Vol. 53, No. 4, Pp.50-58, 2010.

[3] J. Yuan and S.Yu, "Secure and constant cost public cloud storage auditing with de-duplication",in IEEE Conference on Communications and Network Security , Pp.145-153, 2013.

[4] S. Halevi, D. Harnik, B. Pinkas, and A. Shulman-Peleg, "Proofs of ownership in remote storage systems," in Proceedings of the 18th ACM Conference on Computer and Communications Security, Pp. 491-500, 2011.

[5] S. Keelveedhi, M. Bellare, and T. Ristenpart, "Dupless: Server aided encryption for deduplicated storage," in Proceedings of the 22Nd USENIX Conference on Security, Pp. 179-194, 2013.

[6] Renuka Goyal and Navjot Sidhu, "Third Party Auditor: An Integrity Checking Technique for Client Data Security in Cloud Computing", in Proceedings of the (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5, No.3, Pp. 4526-4530, 2014.

[7] S. Nair, Anitha K L and R. Kamala, "Trusted Third Party Authentication in Cloud Computing", in Proceedings of the International Journal of Engineering and Technology (IJERT), Vol. 2 No. 11, Pp. 354-358, 2013.

[8] R. Di Pietro and A. Sorniotti, "Boosting efficiency and security in proof of ownership for de-duplication", in Proceedings of the 7th ACM Symposium on Information, Computer and Communications Security, Pp. 81-82, 2012.

[9] W. K. Ng, Y.Wen, and H. Zhu, "Private data de-duplication protocols in cloud storage", in Proceedings of the 27th Annual ACM Symposium on Applied Computing, Pp.441-446, 2012.

[10] M. Bellare, S. Keelveedhi, and T. Ristenpart, "Message-locked encryption and secure de-duplication," in Proceedings of the Advances in Cryptology { EUROCRYPT, Vol.7881, Pp. 296-312, 2013.

[11] M. Abadi, D. Boneh, I. Mironov, A. Raghunathan, and G. Segev, "Message-locked encryption for lock-dependent messages," in Proceedings of the Advances in Cryptology- CRYPTO,Vol.8042, Pp. 374-391.