IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 4, May 2022
Impact Factor: 6.252

A Secure Data Deduplication in Cloud Using Hashing and AES Algorithm

Mahajan Mayur Pradeep[1], Bhujbal Siddhesh Vasant[2], Abhishek Ganpat Wadhai[3], Patil Sanyukta Madan[4], Prof. Ravindra Borhade[5]
Students, Department of Information Technology[1,2,3,4]
Assistant Professor, Department of Information Technology[1,2,3,4]
Smt. Kashibai Navale College of Engineering, Pune, Maharashtra, India
Savitribai Phule Pune University, Pune, Maharashtra, India

Abstract: *Data de-duplication is a way used to compress data helping withinside the elimination of reproduction copies of data. It has been powerful in cloud garage; it decreases the specified garage area to steady such facts through thinking about de-duplication this concurrent technique has been projected. In the identical way, introducing fewer new replica exams for engineers ought to growth new de-duplication improvement assisting techniques. Security research have proven that this approachment is steady regarding the descriptions proven withinside the projected protection model. This paper will actualize a prototype of a suggested, sanctioned replica test plan and carry out experiments the use of the prototype. The look at will display that the proposed prototype reasons inconsequential overhead-differentiated archetypal processes. This paper offers and analyses a brand new scheme to deal with the difficulty of eternal cloud garage get right of entry to structures through imparting AES Algorithm primarily based totally solution. Addressing the difficulty of Data garage withinside the latest instances is the intention of the paper and is being completed the use of the latest AES Algorithm. Data protection and time constraint for facts retrival from the cloud server is likewise taken under consideration whilst enforcing the scheme.*

Keywords: Deduplication, Storage, Route key, AES Algorithm.

I. INTRODUCTION

As we move towards a more technological era, backing up data to cloud servers is a daily need. Huge amount of data is uploaded to the internet every day, the preservation and security of this data is becoming a concern nowadays Data preservation is very important and in today's commercial era, to preserve This data density, Cloud computing is the most widely used. tools at our discarding. Cloud computing uses networks on distant web-hosted servers to store data, manage data, and process data, using local servers or local machines to store data. 5]. Every cloud storage system has short supply memory and if we start uploading same files to the cloud system, space will be lost and data redundancy will be a big problem. before our eyes [5]. Data deduplication is a proposal used to amend storage. This technique is widely used by many cloud service providers today like Amazon S3, Dropbox, Google Drive [7], Microsoft, etc. Customer Data.But providing valuable information to others is a risky proposition. Researchers are working on this and the best solution they have found is to protect the outsourced data with ciphertext. downloaded, it is decrypted and then displayed to the client. In encryption, data is converted into a known form of ciphertext, but if encryption is done with different keys, it can result in different ciphertexts making deduplication less feasible. . Therefore, encryption is necessary to secure the data [13]. Therefore, deduplication and encryption must work together to ensure the security and reliability of the data.

II. PROPOSED SYSTEM

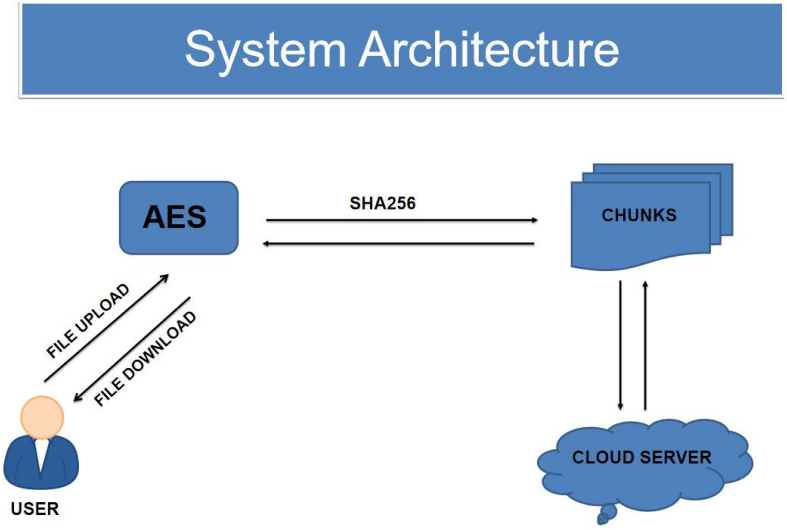
We recommend a value worth based storage substructure that uses encryption established on the AES algorithm, which is an evolving cryptographic innovation to address the challenges of secure information sharing. encryption with a routing key and can taken easy decryption. The key ensures that the shared file is properly stored and that the address is registered. size and type. To enable deduplication and store the transferred information in memory, we use two-way cloud in our cloud infrastructure. A private cloud controls the computer and an open cloud manages capacity. The private cloud comes with a

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 4, May 2022
Impact Factor: 6.252

routing key associated with the comparator ciphertext, which can change the ciphertext for more than one access agreement into equivalent plaintext ciphertexts by several access methods. other without monitoring the hidden plaintext. Once the storage request is received, the private cloud will first verify the validity of the uploaded item through the attached proof.If the proof is valid, the private cloud will run a tag-matching algorithm to see if the same data underneath the ciphertext is stored. It regenerates the ciphertext into an ciphertext of the same plaintext on an access policy that is a set of policies such as a public cloud and a private cloud. The concept of deduplication is efficiency and security achieved through proof of file ownership. Attribute-based encryption of the storage system ciphertext policy and support for secure replication.

III. System Analysis

System architecture is a conceptual model that defines the structure, behavior, and other views of the system. An architectural description is a formal description and representation of a system, organized to support reasoning about the structure and behavior of the system. A system architecture may include system components and developed subsystems that will work together to implement the overall system. We will be using Firebase Realtime Database with Firebase Storage to store our 3D models as well as user information. The presentation layer will contain the activities that the user will interact with and according to their selection logic will be triggered so that the object can be seen in real space.



s

Figure 1: System Architecture

IV. Literature Survey

4.1 Cloud Computing

Cloud computing means storing and accessing data and programs on remote servers hosted on the Internet instead of a computer's hard drive or local server.

4.2 Data Deduplication

This figure shows that data deduplication has evolved as a simple storage optimization technique in secondary storage and then widely adapted in primary storage, as well as storage areas. larger storage area like storage in the cloud. Now, Data Deduplication is widely used by different cloud storage providers like Dropbox, Amazon AWS, Google. Driving, etc. [5]. Once deployed on cloud servers, data exceeds the data owner's secure base. Therefore, most of them prefer to outsource in an encrypted format.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 4, May 2022
Impact Factor: 6.252

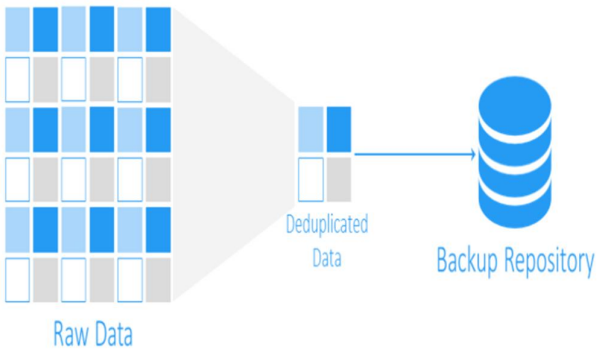


Figure 2: Data Deduduplication

4.3 Types of Data Deduplication

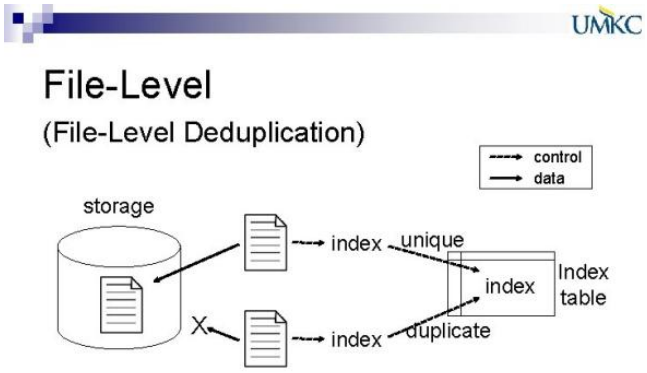


Figure 3: File level data deduplication

File-level deduplication is commonly known as Single Version Archive (SIS) [8]. and update the index; Otherwise, the only repository pointer to an existing file. Thus the same file saves only one instance and then copies all the "originals" instead, while the "original" points to the original file.

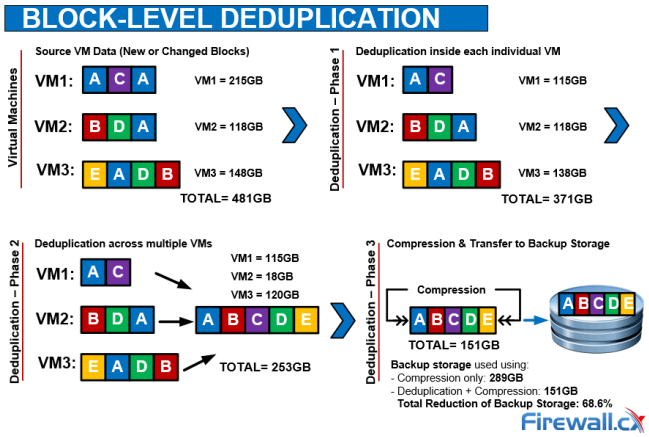


Figure 4: Block level data deduplication

Block-level data deduplication technology to split the data stream into blocks, check the block of data and determine if it encounters the same data before the block. If the block is unique and has been written to disk, its identifier is also stored in the index; Otherwise, the unique drop pointer to store the initial position of the same block of data [9] The hashing

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 4, May 2022
Impact Factor: 6.252

algorithm used to evaluate duplicate data may cause hash error collisions. MD5, SHAI hash algorithm, etc. tested against blocks of data to form a unique code [13]. While there is a possibility of collisions and corruption of hash data, it is much less likely.

V. CONCLUSION

If duplicate data is allowed to be uploaded to the cloud frequently, the cloud storage will be filled with unnecessary and useless data, thus killing our memory and leading to reduced bandwidth and translation. poor customer service. To get around this, we'll use a deduplication function that uses a hash algorithm to deduplicate data. Data must be secure from unauthorized users, cybercriminals. Therefore, encryption is needed, and we will use Advanced Encryption Standard to encrypt the data before uploading it to the cloud.

REFERENCES

[1]. NIST, "NIST Brief Comments on Recent Cryptanalytic Attacks on Secure Hashing Functions and Continued Security Provided by SHA-1", 25th August 2004, http://csrc.nist.gov/groups/ST/toolkit/documents/shs/hash_standards_comments.pdf

[2]. Federal Information Processing Standards Publication 180-3, "SECURE HASH STANDARD", October 2008, http://csrc.nist.gov/publications/fips/fips180-3/fips180-3_final.pdf

[3]. OpenSSL Source Code, http://www.openssl.org/source

[4]. N. Ferguson et al, "The Skein Hash Function Family", Version 1.3, 1st October 2010, http://www.schneier.com/skein1.3.pdf

[5]. N. Jeber, Jalal. (2019). The Future of Cloud Computing Google Drive. 10.13140/RG.2.2.26342.06724.

[6]. Burramukku, Tirapathi & Ramya, U. & Sekhar, M.V.P.. (2016). A comparative study on data deduplication techniques in cloud storage. 8. 18521-18530.

[7]. Ku, Chan-I & Luo, Guo-Heng & Chang, Che-Pin & Yuan, Shyan-Ming. (2013). File Deduplication with Cloud Storage File System. 280-287. 10.1109/CSE.2013.52.

[8]. N. Baracaldo, E. Androulaki, J. Glider, A. Sorniotti, "Reconciling end-to-end confidentiality and data reduction in cloud storage," Proc. ACM Workshop on Cloud Computing Security, pp. 21–32, 2014.

[9]. C. Wang, Z. Qin, J. Peng, and J. Wang, "A novel encryption scheme for data deduplication system," Proc. International Conference on Communications, Circuits and Ssytems (ICCCAS), pp. 265–269, 2010.

[10]. J. R. Douceur, A. Adya, W. J. Bolosky, D. Simon, and M. Theimer, "Reclaiming space from duplicate files in a serverless distributed file system," Proc. International Conference on Distributed Computing Systems (ICDCS), pp. 617–624, 2002.

[11]. D. T. Meyer, and W. J. Bolosky, "A study of practical deduplication," Proc. USENIX Conference on File and Storage Technologies, 2011.

[12]. A. Rahumed, H. C. H. Chen, Y. Tang, P. P. C. Lee, J. C. S. Lui, "A secure cloud backup system with assured deletion and version control," Proc. International Workshop on Security in Cloud Computing, 2011.

[13]. Sheetal U.Jonwal,Pratibha P.Shingar,:"Advanced Encryption standard(AES) implementation on FPGA with hardware in loop", International Conference on Trends in Electronics and Informatics,ICE1 2017,DOI:10.1109/1COEI.2017.8300776.