Secure Deduplication for Cloud Storage by Encrypting Cipher Text Using Aes Algorithm

S.Rama Mani,Akshaya.A.P ,Aiswarya.R ,Ramyasri. M ,Vinisha.J

Abstract: ***Data play a important role in every part of our life. Because processed information is obtained only from the processed data. Those data are gathered from various users and stored in a cloud storage. Cloud is a new technology that is developed to reduce the storage area and cost of storage. Users having same data share a common storage area and that data can be fetched whenever needed. These data's are encrypted and stored as cipher text to avoid data threat. The number of users and the data stored in cloud is increasing exponentially day by day. The only way to reduce the cloud storage is data deduplication, elimination of repeated data in cloud. The data deduplication becomes more and more a necessity for cloud storage providers.To make it more secured the data is encrypted with Advanced Encryption Standard (AES) key.***

Keywords: ***Cloud Computing and cloud platform, Data Privacy, Data Security, Models of Cloud Computing.***

I. INTRODUCTION

Distributed computing is the method for changing how data innovation (IT) is expended and managed,promising improved cost efficiencies, quickened advancement, quicker time-to-showcase, and the capacity to scale applications on interest. Be that as it may, as the distributed computing is rising and growing quickly both adroitly and as a general rule, the legitimate/legally binding, financial, administration quality, interoperability, security and protection issues still posture huge difficulties. In this undertaking, we portray different administration and arrangement models of distributed computing and distinguish real difficulties. Specifically, we talk about three basic difficulties: security and protection issues in distributed computing. A few answers for decrease these difficulties are additionally proposed alongside a concise introduction on the future patterns in distributed computing sending. The utilization of distributed computing has expanded quickly in a considerable lot of the organizations. Cloud registering gives numerous advantages as far as ease and expanding openness of information. Guaranteeing the security of distributed computing is a main consideration in the field of distributed computing condition, as clients frequently store touchy data with distributed storage suppliers yet these suppliers may not be trusted.

Revised Manuscript Received on May 06, 2019
Ms.S.Rama Mani, Assistant Professor,
Akshaya.A.P, Student, Department of Computer Science and Engineering
Aiswarya.R, Student, Department of Computer Science and Engineering
Ramyasri. M, Student, Department of Computer Science and Engineering
Vinisha.J, Student, Department of Computer Science and Engineering

To provide users with advanced security and efficient storage ,the following methods are implemented:
1)Security enhancement by using AES key to encrypt.
2)Deduplication in cloud storage to eliminate the duplicate data.

Existing system

There are many individual existing systems for reducing cloud storage and enhancing the security of the data. The system which requires less storage after deduplication do not offer good security for the data, we have to go for high cost if we need our data to be secured. And the system which offers high security occupies large storage space, there are many repeated data present[1]. Traditionally some of problems are there in the existing system. Customers must be wary while using drag/drop to move a report into the circulated stockpiling envelope, This will everlastingly move your record from its appropriated stockpiling territory.

Accessibility: If you have no web affiliation, you have no passage to your information. Information's are not confirmed because of one measurement encryption. Any development that prompts data incident/debasement through to interference of average business assignments is called as threat.

Proposed system

In this system the security and low storage comes together at low cost, because we enhance the security by implementing high level security algorithm and a efficient storage service to reduce the storage. So, This will produce following features

- ✓ Make the data more secured.
- ✓ Reduces storage space.
- ✓ Reduces storage cost.

II. MATERIALS AND METHODS

Cloud storage

Clients are furnished with benefits, extending from cost sparing and simplified comfort, to versatility openings and adaptable administration in distributed storage. These incredible highlights are pulled in by an ever increasing number of clients to use and store their own information to the distributed storage[3].

As indicated by the investigation report, the volume of information is relied upon to accomplish 50 trillion gigabytes in 2020. Distributed computing server farms are demonstrated upon a basic plan that is for disappointment framework. They utilize minimal effort, reason fabricated, versatile arrangements, that additionally incorporates servers, stockpiling frameworks and systems administration items, while as yet using standard conveyance models and gigantic economies of scale. Distributed computing server farms, notwithstanding, don't mirror the rack frameworks intended for the conventional mass-IT showcase[4]. Distributed computing is generally depicted in any of the two different ways. It is either founded on the sending model, or on the administration that the cloud is putting forth. In light of an administration that the cloud show is putting forth, we are talking about either:IaaS- Infrastructure-as-a-Service or PaaS- Platform-as-a-Service or SaaS- Software-as-a-Service or Storage-as-a-service

Encryption

In cryptography, encryption is the way toward encoding message or data so that just approved gatherings can get to it and the individuals who are not approved can't. Encryption does not itself avert impedance, however denies comprehensible substance to would-be interceptor. In an encryption plot, the proposed data or message, alluded to as plaintext, is encoded utilizing an encryption calculation producing figure - content that can be perused just whenever decoded[5]. For specialized reasons, an encryption plot as a rule utilizes the pseudo-arbitrary encryption key which is produced by calculation. Subsequently it is important. Some propelled encryption calculations which have been connected into distributed computing decreases the security issues. In a training called crypto-destroying, the keys utilized for encryption can basically be erased when there is no more utilization of the information.

Attribute-based encryption (ABE)

Trademark based encryption is one of the sort of open key encryption in which the puzzle key of a customer and the figure content are dependent upon attributes (for instance the country in which he lives, or subject to the kind of enrollment he has). In such a structure, the deciphering of a figure content is possible if and just if the course of action of properties of the customer key matches the attributes of the figure content.

Cipher text policy ABE (CP-ABE)

In the CP-ABE, the encryptor has a command over access procedure. The primary research work of CP-ABE is chiefly centered around the plan of the entrance structure.

Decryption

In cryptography, unscrambling is the way toward disentangling a message or data so that just approved gatherings can peruse the encoded data and the individuals who don't have a clue about the way to unravel the scrambled information they can't peruse the verified data. Distributed storage contains the scrambled information so it is in charge of both encryption and decoding process. Putting away the information in encoded structure is a typical strategy for data security assurance thus we did it twice in our task.

Deduplication

In registering, where deduplication is a specific information pressure system for dispensing with copy duplicates of rehashing information. This strategy is utilized to improve capacity usage and can likewise be connected to organize information which exchanges and lessens the quantity of bytes that must be sent[6]. In the deduplication procedure, interesting information, or byte designs, are distinguished and are put away amid a procedure of investigation for sometime later.As the examination proceeds, different pieces of information are contrasted with the put away duplicate and at whatever point a match happens, the excess information is supplanted with a little reference that focuses to the put away information. Given that a similar byte example may happen handfuls, hundreds, or even a large number of times, the measure of information that must be put away or exchanged can be incredibly diminished and deduplication is essentially pressure procedure for evacuating repetitive information that additionally clarifies the deduplication procedure before putting away information onto memory. Deduplication can be ordered as record level deduplication and the square dimension deduplication dependent on granularity[7]. Record level deduplication considers the whole document, along these lines even little update or add makes the record not the same as past adaptation of it and in this way decreasing deduplication proportion. Where as if there should be an occurrence of square dimension deduplication information squares are considered for the deduplication. Deduplication can additionally arranged dependent on area of deduplication that is recorded as 1) Client side of deduplication and as 2) Source side deduplication.

Advantages of cloud storage

Accessibility: Files in the cloud can be gotten to with an Internet association from anyplace on the planet. This gives access as well as enables you to move past time zone and geographic area issues. Lower Operation Cost: Distributed storage for your business will come at almost no expense for a little or medium-sized association. This will decrease your yearly working expenses and considerably more funds as it doesn't rely upon inward capacity to store data remotely.Speed: In cloud storage, you only pay for the amount of storage you require. on the off chance that your business encounters an appropriate development, at that point the cloud administrator can assist you with accommodating the comparing development in information stockpiling needs. The main thing you need to do is shift the amount you pay to broaden the capacity you have. This is additionally relevant regardless of whether your business therapists and you require less extra room at a diminished rate which is more verified than executing the old joined encryption system. Deduplication results in the involving low stockpiling in the cloud.

Disadvantages of cloud storage

Security and protection in the cloud: The primary concern is with significant and vital information being put away remotely. Embracing cloud innovation, gives touchy business data to an outsider cloud specialist co-op and this could conceivably put your organization in danger. Along these lines, it is critical to pick a dependable specialist organization that you are sure will keep your data secure.

Transmission capacity restrictions: Based on what administration you pick, there could possibly be a transfer speed remittance. In the event that your business outperforms the remittance, at that point charges could be exorbitant relying upon that.A few sellers give boundless transmission capacity and this is something to contemplate while picking the correct supplier.Powerlessness to Attacks: With your business data put away in the cloud, there is a defenselessness notwithstanding for the outer hacking assaults[8]. As the web isn't totally verify, and therefore, there is dependably the likelihood of stealth of delicate information.

Information Management: Managing information in the cloud can be an issue since distributed storage frameworks have their very own decided structures. The current stockpiling the executives arrangement of your business may not generally incorporate well with the cloud seller's framework by making it a major breakdown.

Lifetime costs: With open distributed storage, the cost expenses throughout the years may increment and will in general include as no one will give free administration over years. This is equivalent to purchasing another vehicle with a substantial forthright expense. The comfort of rent installments may look engaging toward the start yet anyway you will owe for mileage overage and need to pay a ton to keep the vehicle. This is the point at which the lifetime costs will hit you at last. In the event that your applications are nearby and your information is in the cloud, at that point it can even add to the systems administration costs[9].

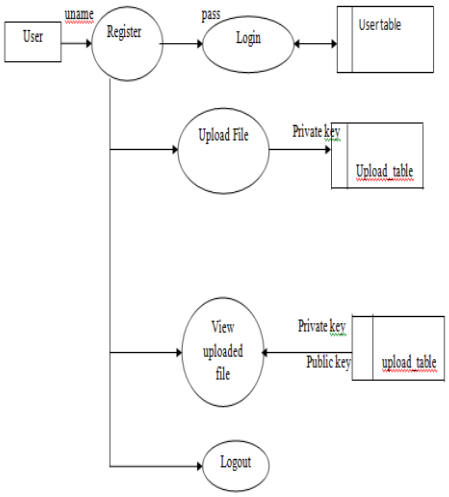
Compliance:Depending on the level of regulation within your organization, it may not be possible to work within the public cloud. This is especially the case for healthcare, financial services and publicly traded companies that have to be very careful when considering this option.

III. METHODOLOGIES

Property based encryption (ABE) is a sort of calculation of open key cryptography in which the private key is utilized to decode information relies upon certain client traits, for example, position, spot of living arrangement, kind of record. Trait based encryption (ABE) accomplishes the two information security and access control. It is finished by conceding diverse unscrambling rights to clients dependent on qualities, for example, the client's specialization and position. Clients or their qualities must be repudiated as needs be so the denied clients can never again unscramble information since clients and their traits change after some time inside the framework. In the Implementations of ABE, mixture encryption is utilized for productivity information is encoded with AES[10]. And furthermore AES key is scrambled with ABE. The information must be re-scrambled with another AES key so that renounced clients can't unscramble information utilizing the old AES key that they may have acquired before disavowal. Since renouncement happens just on the ABE ciphertext this re-encryption is finished.Re-encryption on the information ought to be done to abstain from unscrambling information. There are two strategies for renouncing clients or their traits in ABE. The two techniques are disavowal rundown and intermediary re-encryption. Renouncement list is utilized to deny clients by encoding information. This encryption is with a rundown of clients to such an extent that clients having a place with the rundown can't unscramble information. Intermediary re-encryption is utilized for repudiating ascribes to refresh ciphertexts. It is likewise utilized for disavowing ascribes to refresh non-repudiated clients' keys however clients with a denied property can never again utilize their key to decode the refreshed ciphertext.Intermediary re-encryption strategy receives ABE in half and half with a symmetric encryption conspire rather than AES.

IV. BLOCK DIAGRAM

Fig:1-Level 1 diagram



V. IMPLEMENTATION

User Registration:

User Registration module allows public users site to register and access their content. You can use the module to register users for other custom modules that support personalization and user specific handling. User can login by giving their registered user_name and password. Only the authenticated user can enter to access the contents.

Upload file:

In this module the registered users can upload their files into the storage. This module involves the first step of selecting the file to be uploaded and classifying the file to find its class. The class of the file is based on the sensitivity of the file contents.

The high class file with more sensitive and highly confidential elements gets stored under cloud service provider 1. Similarly, the medium and low class file get stored under cloud service provider 2 and cloud service provide 3 respectively. The second step of this module is to encrypt the file, user can provide their own encryption key and encrypt it using AES algorithm. Now this encrypted file is sent to the cloud admin for further process.

Admin module:

There are 3 cloud service providers namely csp1, csp2 and csp3 based on their sensitivity of the data stored. The second level of encryption (encrypting cipher text) happens in this module, when the file is assigned to their respective storage the admin encrypts the file with the homographic key using AES algorithm. This double encrypted file is now moved to the cloud storage.

Access files:

In this module, the user can access their stored files. For this the user have to decrypt the file by providing the user encryption key and homographic key. The homographic key is attached with the file name. Now the decrypted file can be accessed by the user. The double encryption makes the file to be stored securely.

VI. CONCLUSION

In this framework, we scramble the information present in the cloud twice by utilizing intermediary re-encryption technique. This is recommended that utilizes ABE in half breed with symmetric intermediary re-encryption conspire. So information can be re-encoded by cloud servers. The information proprietor just needs to create a lot of re-encryption keys and ABE figure content scrambling the new keys then send both to the cloud for re-encryption. Since the one and only way is the correspondence with little measure of information is important, the correspondence cost is diminished contrasted with the minor arrangement.

REFERENCES

1. "A Hybrid Cloud Approach for Secure Authorized Deduplication" by Jin Li, Yan Kit Li, Xiaofeng Chen, Patrick P.C. Lee, and Wenjing Lou in 2015.
2. "Secure Deduplication for Cloud Storage Using Interactive Message-Locked Encryption with Convergent Encryption, To Reduce Storage Space" by Jayapandian N,Md Zubair Rahman A M J in 2018.
3. "Secure Auditing and Deduplicating Data in Cloud" by Jingwei Li, Jin Li, Dongqing Xie, and Zhang Cai in 2106
4. A. Sahai, and B. Waters for "Fuzzy identity-based encryption,"Advances in Cryptology–EUROCRYPT 2005, LNCS, vol. 3494, pp.457-473, 2005.
5. Y. Cheng et al for "Efficient revocation in ciphertext policy attribute based encryption based cryptographic cloud storage," Journal of Zhejiang University SCIENCE C, vol. 14, Issue 2, pp. 85-97, 2013.
6. F. Wang, J. Mickens, N. Zeldovich, V.Vaikuntanathan,"Sieve:Cryptographically enforced access control for user data in untrusted clouds," Proc. of 13th USENIX Symposium on Networked Systems Design and Implementation means (NSDI '16), pp. 611-626, 2016.
7. D. Boneh, K. Lewi, H. Montgomery, and A. Raghunathan, "Keyhomomorphic PRFs and their applications," Advances in Cryptology-CRYPTO 2013, LNCS, vol. 8042, pp. 410-428, 2013.
8. S. Myers and A. Shull, "Efficient hybrid proxy re-encryption for practical revocation and key rotation," Cryptology ePrint Archive,https://eprint.iacr.org/2017/833.
9. A. Syalim, T. Nishide, and K. Sakurai for "Realizing proxy re-encryption in the symmetric world," Proc. of Int'l Conf. onInformatics Engineering and Information Science, CCIS, vol. 251, pp.259-274, 2011.
10. N.Attrapadung and H. Imai for "Conjunctive broadcast and attribute-based encryption," Pairing-Based Cryptography–Pairing 2009