International Journal of Advance Research in Science and Engineering
Volume No. 11, Issue No. 09, September 2022
www.ijarse.com
IJARSE
ISSN 2319 - 8354

Secure and Efficient Data Retrieval in Cloud Computing using Map Reduce

[1]Sonal, [2]Prof. Harshita

[1]M.Tech Scholar, [2]Assistant Professor

Department of Computer Science Engineering

Neelkanth Institute of Technology, Meerut, India

Abstract

Cloud computing is a remarkable technology which provide solution for a defended, on-demand, and dynamically scalable computing infrastructure towards many applications. their quality services, including server, database, networking, software etc, to its users such as individuals, industries as well as organization, are attracting towards the cloud to store their important, valuable, confidential data and get easy to fetch data over the internet anywhere anytime during the execution time when number of node increase the probability of failure increase fastly since its possibilities are less to prevent failure but in Map Reduce is a programming system for distributed processing large-scale data in an efficient and effective manner on a private, public, or hybrid cloud. The developed system is implemented using the El-Gamal cryptography algorithm to provide security through effective key generation techniques and encryption strategy over a cloud storage environment.

Keywords: Cloud Computing, Cloud, MapReduce, DataMigration., Data Sharing, Security, Privacy, EL-Gamal.

Cloud Computing

Cloud Computing [1] provides services to its users to upload their data safely in less duration on cloud servers and access that data anytime, everywhere through the network. Cloud computing offers different service models to its user such as SaaS (software as a service), IaaS (infrastructure as a service), PaaS (platform as a service), STaaS (storage as a service), SECaaS (Security as a service) & many more.In cloud services commonly user use storage which provides many advantages, client use it to store their files on cloud to avoid the disruption of storing , upholding the data files locally and data can be access from any geographical location and reduces the maintenance of hardware and software. However, since the stored data is on cloud server i.e. at a remote location, how to get the verification about the stored data. Since the cloud users do not able to have physical check over outsourced data, that makes data integrity checking in cloud environment a substantial job. One of the major challenges in cloud storage[2][3][4] service is cloud data integrity verification. One easy way is to load the entire data files on local system and carry out the integrity checking. However, this results

in severe I/O overhead on the server by which transmitting the entire data file over the network increase the network traffic by which overloading is a major task by which it difficult to identify the damage data file so, it become too late to recovery ,while accessing the store data. Therefore to assure data authenticate and authorization, it is necessary to introduce an effective method for clients to validate the authenticity of the data stored on the cloud. To completely guarantee on cloud user's data integrity, it is more significant to allow public auditing service for client's outsourced data. Public auditing service makes use of an auditor, usually a Third Party Auditor (TPA)where data owner uploaded files on cloud which to be frequently audit . These TPAs possess knowledge and expertise that clients do not and are allowed to check the integrity of client's outsourced files on cloud when needed. Thus the Third Party Auditing mechanism provides an efficient solution for cloud users to verify their data storage correctness on cloud at anytime. The service providers can also gain valuable insights from the audit results provided by these TPAs which can further help to improve their cloud service over the internet.

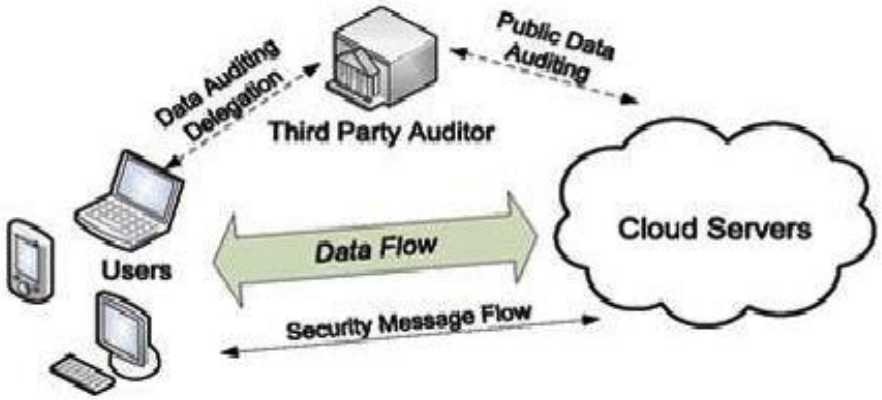


Fig 1: Cloud Storage Architecture

In public auditing, TPA will inquire the CSPs to prove that data files uploaded by a particular cloud user are safe and unmodified. However through this process, the original data gets revealed to Third Party auditors. TPA must not be permitted to access original data contents for security and privacy reasons. Thus encrypting the original data files before integrity verification is also necessary. The proposed system makes use of public key based ElGamal encryption scheme to improve data storage security in cloud for its user.

Map Reduce

MapReduce and HDFS are the two major components of Hadoop which makes it so powerful,effective and efficient to use. MapReduce is a programming model used in distribution manner over a large data whose processing is done parallel.[5] The data is first split and then

combined to produce the final result. The libraries for MapReduce is written in number of programming languages with various different-different optimizations result. The purpose of MapReduce in Hadoop is to Map each of the jobs and parcels out work to various nodes within the cluster network or map and to reduce the result from each node into a cohesive answer.The MapReduce task is mainly divided into two phases Map Phase and Reduce Phase.[6]

In **MapReduce**, we have a client. The client will submit the a set of data(job) of a particular size to the Hadoop MapReduce Master. Now, the MapReduce master will divide this set of Data where individual Data (job) are broken down into job parts (tuples/key/value pair).[7] These job-parts are then made available for the Map and Reduce Task. This Map and Reduce task will contain the program as per the requirement of the use-case that the particular company is solving. The developer writes their logic to fulfill the requirement to particular that the industry requires. The input data which we are using is then fed to the Map Task and the Map will generate intermediate key-value /tuples to pair as its output. [8]The output of Map i.e. scaling the application to run over no of machine in a client is merely a configuration change what attract many programmer to use the map reduce model.

El gamal Algorithm

The ElGamal cryptosystem[9] is based on the difficulty of discrete logarithm problem which implies that the discrete logarithms are extremely difficult to compute in defined amount of time, whereas the inverse operations of the power are easy to compute. ElGamal is a public encryption algorithm in which use of a random exponent k. This k is used in place of private exponent of receiver.[10] Thus the entire operation is performed by one party i.e. the party by which encryption is done[11]. Thus the encryption can be performed in one direction, without interruption of the second participant. Following are the steps involved in ElGamal encryption algorithm:[12][13]

A. Key Generation

The key generation process works as follows:

a. Assume a large prime number p.

b. Choose a primitive element g modulo p.

c. Choose a private key a randomly from {1, ..., p-1}.

d. Compute public key y as follows: a. y = ga mod p

B. Encryption The encryption algorithm is as follows:

The plaintext is expressed as a set of numbers modulo p.

Data owner encrypts a message M, CP be the cipher text; CP comprises of two values ciphertext1 (y1) and ciphertext2 (y2).

a. Generate a random number k less than p

b. Compute two values y1 and y2 , where

y1 = gk mod p

y2 = M xor yk

c. Transmit the cipher text CP consisting two values y1 and y2.

C. Decryption

Upon receiving the cipher text CT (y1 and y2), the receiver computes original message M as:

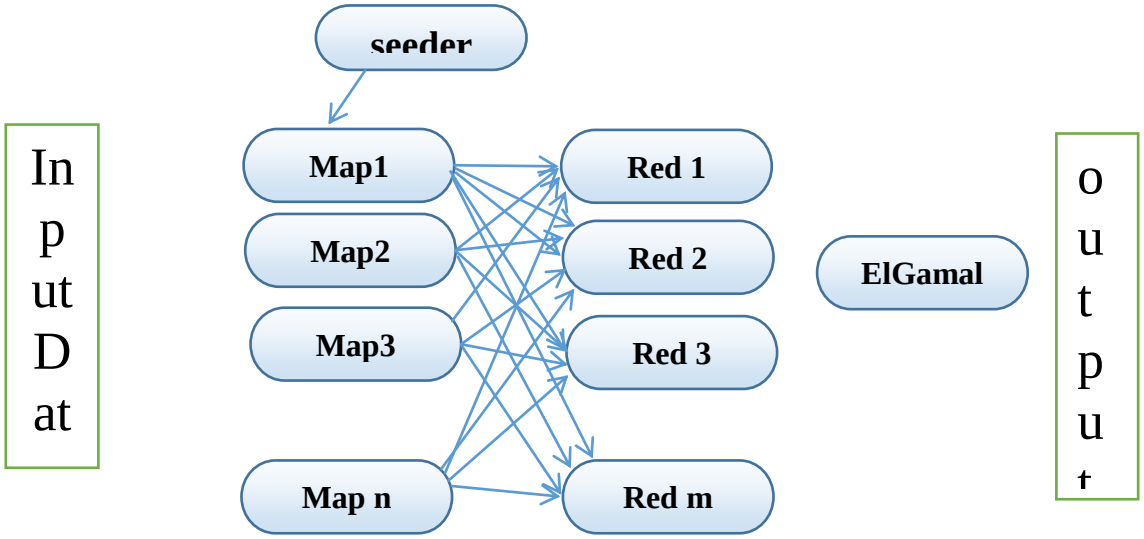
M = (y1 a mod p) xor y2.

Related work

Cloud Computing is a range of services delivered over the internet or the cloud to its user. Transporting computing power (CPU, RAM, Network Speeds, Storage OS software)by using remote server to store & access data due to Cloud Computing. Data in the Cloud can be in various states: at-rest, in-use, in-transit. The data do not have the same level of security requirements at the time. Data being processed can not be protected similar as data intransit or at rest. In a network, there is no such complete security solution to secure user data and app, or services, but through risk management reduction in the level of risks can be done . Map Reduce is a programming system for distributed processing large-scale data in an efficient and effective manner on a private, public, or hybrid cloud. In this paper, we investigate and discuss security and privacy challenges and requirements, considering a variety of adversarial capabilities, and characteristics in the scope of Map Reduce through Elgamal algorithm.[14][15]

The previous work was comparison between the parallel method and the sequential method. Several security issues have been identified in the current Cloud Era [16]. Data is not safe at the time of fetching data from different servers/security for distributed cloud computing. A secure and non faulty migration technique is necessarily required that can efficiently transfer the data between the servers using Map Reduce based algorithm. Which helps in fetching the data with minimum data loss and is more secure manner by encrpyting the data through Elgamal algorithm.

International Journal of Advance Research in Science and Engineering
Volume No. 11, Issue No. 09, September 2022
www.ijarse.com
IJARSE
ISSN 2319 - 8354



Proposed work

During retrial of data from cloud environment it is very essential to secure our data while we are fetching from hidden web like shopping site data like amazon, mail data (enterprise), enterprise storage. To ensure the protection of data from such cloud storages we need an encryption mechanism after the Maping process in MapReduce environment. The following steps has to be keep in mind while creating algorithm for secure MapReduce data retrial mechanism.

Step1:

A Seeder/ Job descriptor process receives a job descriptor, which specifies the MapReduce job to be executed. The job descriptor contains, among other information, the location of the input data, which may be 6accessed using a cloud storage.[17][18]

Step2:

According to the job descriptor, the Seeder starts a number of mapper and reducer processes on different cloud storage machines. At the same time, it starts a process that reads the input data from its location, partitions that data into a set of splits, and distributes those splits into various mappers.

Step 3:

After receiving its data partition, each mapper process executes the *map* function (provided as part of the job descriptor) to generate a list of intermediate key/value pairs. Then these pairs are grouped on the basis of their keys.

Step4

All pairs with the same keys are assigned to the same reducer process. Hence, each reducer process executes the *reduce* function (defined by the job descriptor), which merges all the values associated with the same key to generate a possibly smaller set of values.

Step5:

All the data retrieved after reduce function must be encrypted by **El gamal Algorithm**.

Step 6:

Then results generated by each reducer process is collected and delivered to a location specified by the job descriptor, so as to form the final output data. The final data will be Secure and effectively processed from cloud storage environment.[19]

Conclusion

Information secure with complete protection is thesignificant worries for users while cloud computing.[20] Specifically, applying security concern for multiple users and furthermore ensuring their data protection turns into a difficult task. In this paper, a secure groupafterwards they are merge in the encrypted text, Elgamal techniques is used. Each data file's index is combined into a single index. This is a secure pursuit convention that enables several data owners to encode files and indexes using separate keys. The cloud server can then combine encoded indexes without knowing any data other current techniques for keyword mapping are less efficient. From the security purpose, we have proposed secure Elgamal algorithm with fulfill the distribute large data in form of large scale of data in the form of job which are mapped through reducer security requirement. Again efficiency of our system depends on the proposed concept of an Elgamal algorithm search strategy with required less[21] computation time .

Future work

In our scheme, the data owner could encrypt his private data and share them with a group of data access gadgets at the same time helpfully dependent on the proposed procedure. To create mapping to the input data which Iis slipt into number of job by reduction is made to particular job afterwards they are merge in the encrypted text, Elgamal techniques is used. In futire work we will compare this mechanism over different cloud storage environment to find the efficiency and behaviour of our method of secure data retrial over cloud storage using MapReduce mechanism.

References

1. Y. Miao, X. Liu, K.-K. R. Choo, R. H. Deng, J. Li, H. Li, and J. Ma, "Privacy-preserving attribute-based keyword search in shared multi-ownersetting," IEEE Transactions on Dependable and Secure Computing,2019.

2. D. Wu, J. Yan, H. Wang, D. Wu, and R. Wang, "Social attribute aware incentive mechanism for device-to-device video distribution," IEEE Transactions on Multimedia, vol. 19, no. 8, pp. 1908–1920, 2017.

3. D. Wu, Q. Liu, H. Wang, D. Wu, and R. Wang, "Socially aware energy-effificient mobile edge collaboration for video distribution," IEEE Transactions on Multimedia, vol. 19, no. 10, pp. 2197–2209, 2017.

4. D. Wu, S. Si, S. Wu, and R. Wang, "Dynamic trust relationships aware data privacy protection in mobile crowd-sensing," IEEE Internet of Things Journal, vol. PP, no. 99, pp. 1–1, 2017.

5. Sherif Sakr, Anna Liu, Ayman G. Fayoumi, "The Family of MapReduce and Large Scale Data Processing Systems" arXiv:1302.2966v1 doi: http://arxiv.org/pdf/1302.2966.pdf 13 Feb 2013

6. Jeffrey Dean and Sanjay Ghemawat, MapReduce: Simplied Data Processing on Large Clusters, Google, Inc., OSDI ,pages 1-13, 2004

7. Article by IBM - "What is MapReduce?"www01.ibm.com/software/data/infosphere/hadoop/mapreduce/"

8. Tyson Condie, Neil Conway, Peter Alvaro, Joseph M.Hellerstein, Khaled Elemeleegy, Russel Sears, Map Reduce Online, pages 1-14 , Oct 9,2009

9. ElGamal Cryptosystem, http://lxmayr1 .informatik. tumuenchen. de/konferenzen/ Jass05/courses/ 1/papers/ meier_paper.pdf

10.Taher ElGamal (1985). "A Public-Key Cryptosystem and a Signature Scheme Based on Discrete Logarithms" (PDF). IEEE Transactions on Information Theory. **31** (4): 469–472. CiteSeerX 10.1.1.476.4791. doi:10.1109/TIT.1985.1057074. (conference version appeared in CRYPTO'84, pp. 10–18)

11.Mike Rosulek (2008-12-13). "Elgamal encryption scheme". University of Illinois at Urbana-Champaign. Archived from the original on 2016-07-22.

12.H.Ong and C. Schnorr, Signatures Through Approximate Representations by Quadratic Forms, to be published.

13.H.Ong, C. Schnorr, and A. Shamir, An eflicient Signature Scheme Based on Quadratic Forms, pp 208-216 in Proceedings of 16th ACM Symposium on Theoretical Computer Science, 1984

14.R. Rivest, A. Shamir, and L Adleman, A Method for Obtaining Digital Signatures and Public Key Cryptosysterns. Communications of the ACM. Feb 1978, volume 21. number 2. 120- 128.

15D.Coppersmith. Fast Evaluation of Logarithms in Fields of Characteristic two, IEEE Transactions on Information Theory IT-30 (1964), 587-594.

16. Sonal and Vijay, "A Study on Map Reduce Based Secure and Fault Tolerance Data Migration in Cloud Computing"

17.L. Adleman, A Subexponential Algorithm for the Discrete Logarithm Problem with Applications to Cryptography, Proc. 20th IEEE Foundations of Computer Science Symposium (1979). 55-50.

18. Sreenivasa Rao Veeranki, "The Method for Pharma Growth Throughout The Drug Lifecycle With Artificial Intelligence In Life Science, International e-Conference on Innovation and Emerging Trends in Engineering, Science and Management, ISBN: 978-93-91535-38-4, 24th and 25th June 2022

19 T. E. Gamal, "A public key cryptosystem and a signature scheme based on discrete logarithms," IEEE Trans. Information Theory, vol. 31, no.

4, pp. 469–472, 1985. http://dx.doi.org/10.1109/TIT.1985.1057074

20. C. Hazay, G. L. Mikkelsen, T. Rabin, and T. Toft, "Efficient RSA key generation and threshold paillier in the two-party setting," in Topics in Cryptology - CT-RSA 2012 - The Cryptographers' Track at the RSA Conference 2012, San Francisco, CA, USA, February 27 - March 2, 2012. Proceedings, 2012, pp. 313–331. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-27954-6-20

21. I. Damg ˚ard, M. Jurik, and J. B. Nielsen, "A generalization of paillier's public-key system with applications to electronic voting," Int. J. Inf. Sec., vol. 9, no. 6, pp. 371–385, 2010. [Online]. Available: http://dx.doi.org/10.1007/s10207-010-0119-9

22. Veeranki, S. R., & Varshney, M. (2022). Comparative analysis of thyroid disease and predict them using machine learning techniques. International Journal of Health Sciences, 6(S3), 11005 – 11014. https://doi.org/10.53730/ijhs.v6nS3.8459

23 Veeranki, S. R., & Varshney, M. (2022). Application of data science and bioinformatics in healthcare technologies. International Journal of Health Sciences, 6(S4), 5394 – 5404. https://doi.org/10.53730/ijhs.v6nS4.10728

24. Sreenivasa Rao Veeranki, Manish Varshney. (2022). Trends and Application of Data Science in Bioinformatics. Design Engineering, (1), 3541 - 3555. Retrieved from http://www.thedesignengineering.com/index.php/DE/article/view/950

25. Sreenivasa Rao Veeranki, "Metagenomics and Single-Cell Technologies for Microbiome Big-Data Mining: Precision Medicine in the Making" , International Journal of Mechanical Engineering, https://kalaharijournals.com/journals.php,ISSN: 0974-5823 Vol. 7 (Special Issue, Jan.-Mar. 2022)

26. Sreenivasa Rao Veeranki and Manish Varshney, " Intelligent Techniques and Comparative Performance Analysis of Liver Disease Prediction " , International Journal of Mechanical Engineering , https://kalaharijournals.com/ijme-vol7-issue-jan2022part2.php , ISSN: 0974-5823 Vol. 7 No. 1 January, 2022,

27. Sreenivasa Rao Veeranki, “Bioinformatics and Data Science in Industrial Microbiome Applications: A Review”, ISSN: 0974-5823 Vol. 7 (Special Issue 5, April-May 2022) International Journal of Mechanical Engineering, https://kalaharijournals.com/journals.php

28. Sreenivasa Rao Veeranki.” ARTIFICIAL INTELLIGENCE FACILITATED IN BUSINESS INTELLIGENCE ” ,International Journal of Management Technology and Engineering,http://ijamtes.org/VOL-10-ISSUE-3-2020/, Sreenivasa Rao Veeranki - Maharishi university of Information Technology, Lucknow, INDIA.Page No : 446-456 , DOI:16.10089.IJMTE.2020.V10I03.20.3587,ISSN NO: 2249-7455

29. Sreenivasa Rao Veeranki .” ARTIFICIAL INTELLIGENCE IN HANDLING NATURAL SCIENCES ” . ” ,International Journal of Management Technology and Engineering,http://ijamtes.org/VOL-10-ISSUE-4-2020/, Sreenivasa Rao Veeranki - Maharishi university of Information Technology, Lucknow, INDIA. Volume X, Issue IV, APRIL – 2020, Page No : 518-528,DOI:16.10089.IJMTE.2020.V10I04.20.3665, ISSN NO: 2249-7455

30. Sreenivasa Rao Veeranki, Manish Varshney, "Role of Bioinformatics In Biotechnology Concerning AI", IJRAR - International Journal of Research and Analytical Reviews (IJRAR), E-ISSN 2348-1269, P- ISSN 2349-5138, Volume.6, Issue 2, Page No pp.8-17, June 2019, Available at : http://www.ijrar.org/IJRAR19K9397.pdf

31. Sreenivasa Rao Veeranki, Manish Varshney,"Bioinformatics and Data Science in Medical Research", International Journal of Emerging Technologies and Innovative Research (www.jetir.org), ISSN:2349-5162, Vol.6, Issue 1, page no.204-214, January-2019, Available :http://www.jetir.org/papers/JETIR1901G29.pdf

32. Sreenivasa Rao Veeranki, “Revolutionary Transform of Supply Chain Design & Management Using Artificial Intelligence and Bigdata Analytics” , International Conference on Recent Development in Engineering Sciences, Humanities and Management, ISBN : 978-81-943584-9-7, 24th-25th January 2020

33. P. Paillier, “Public-key cryptosystems based on composite degree residuosity classes,” in EUROCRYPT, 1999, pp. 223–238. [2] X. Yi, A. Bouguettaya, D. Georgakopoulos, A. Song, and J. Willemson, “Privacy protection for wireless medical sensor data,” IEEE Trans. Dependable Sec. Comput., vol. 13, no. 3, pp. 369–380, 2016