SEPDP: Secure and Efficient Privacy Preserving Provable Data Possession in Cloud Storage

Mr.Sivasakthi.C
dept.Computer science
Mangayarkarasi College of engineering
Madurai,India

Mr.Nantha kumar.K
dept.Computer science
Mangayarkarasi College of engineering
Madurai,India

Mr.Ajith.S
dept.Computer science
Mangayarkarasi College of engineering
Madurai,India

Mrs.Karpagam.M
dept.Computer science(Assistant professor
Mangayarkarasi College of engineering
Madurai,India

***Abstract*—Cloud computing is an emergent paradigm to give dependable and versatile foundation empowering the clients to store their data and the information purchasers can access the data from cloud servers.This worldview decreases stockpiling and support cost of the information proprietor. At the meantime, the information proprietor loses the physical control and ownership of data which leads to many security dangers. Therefore, auditing service to check data integrity in the cloud is essential. This issue has become a challenge as the possession of data needs to be verified while maintaining the protection. To address these issues this work proposes a protected and effective security saving provable information ownership . Further, we stretch SEPDP to support different proprietors, data dynamics and clump verification. The most alluring e feature of this scheme is that the reviewer can verify the possession of data with low computational overhead.**

***Index Terms*—Cloud Security,preserving,block rule.**

I. Introduction

Storage-as-a-service has emerged as a commercial alternative for local data storage due to its characteristics include less initial infrastructure setup, relief from maintenance overhead and universal access to the data irrespective of location and device. Though it provides several benefits like cost saving, accessibility, usability, syncing and sharing, it raises several security threats as data is under the control of the cloud service provider (CSP). CSP can discard the rarely accessed data to save space and earn more profit, or it can lie about the data loss and data corruption, as a result of software/ hardware failure to protect its reputation. Therefore, it is necessary to check the possession of data in the cloud storage data generated from the wearable devices has high sampling rate and hence, it needs to be stored and handled carefully at the cloud centric data server. A wearable sensor based medical system includes various flexible sensors worn on various parts of the body of a person (patient), including into textile fiber, clothes, elastic bands or even these can be directly Map reduce, great progress has also been made with hardware. Nowadays it is common for commodity clusters to have processors of more and more in-chip cores (referred to as many-core cluster hereafter. attached to the human body in case the devices are implantable medical devices. Traditional cryptographic solutions for integrity checking of data, either need a local copy of the data (which the data users (DUs) do not have) or allow the DUs to downloads the entire data. Neither of these solutions seems practical as earlier one requires extra storage and later alternative increases the file transfer cost. To address this issue, several schemes including are proposed which employ block less verification to verify the integrity without downloading the entire data. One of the attractive features of these works is to allow the public verifier to verify. With public auditability, DUs can recourse the auditing task to a third party auditor (TPA). It has expertise and capabilities to convince both the CSP These schemes use provable data possession (PDP) technique, which gives probabilistic data possession guarantee by randomly verifying few blocks for ensuring possession of data in the untrusted cloud storage. Privacy preserving is essential to prevent TPA to infer the data using the cloud servers response while auditing. However, the schemes proposed do not achieve privacy preserving requirement. Though data dynamics is an important feature to facilitate the data owners to insert, modify, and delete on a particular block of data, without changing the meta-data of other blocks, the techniques proposed in do not achieve data dynamics requirement. Meanwhile, the schemes like could not achieve batch auditing requirement which ensures that TPA should be capable enough to deal with the multiple numbers of simultaneous verification requests from different DUs. This property is to save computation and communication

Identify applicable funding agency here. If none, delete this.

cost between CSP and TPA. Unfortunately, the schemes use pairing based cryptographic operations which are intensive computation and need more time. We propose a secure and efficient privacy preserving provable data possession scheme (SEPDP) for cloud storage. It operates in three phases, namely, key generation, signature generation and auditing phase. Most attractive feature of SEPDP is that it does not use any intensive computation like pairing based operation. Further, we extend SEPDP to support multiple data owners, batch auditing, and dynamic data operations. A probabilistic analysis to detect the integrity of the blocks stored at CSP. We evaluated the performance of the proposed scheme and compared with some of the existing popular mechanisms. We observe that the total time for verification carried out by TPA in the proposed scheme is less than that of the existing schemes. This signifies that SEPDP is efficient and suitable to implement the verification at the low powered devices. Remote data integrity checking protocols can be broadly categorized into two kinds. The deterministic guarantee based, verify each block of data and therefore require a significant amount of storage and computation. Alternative kind of schemes called provable data possession (PDP) include , use probabilistic checking method, in which a few blocks are randomly selected to detect manipulation. PDP is introduced in, that uses random sampling of a few blocks for integrity verification. Two different integrity verification mechanisms. One uses pseudo-random function (PRF) which fails to provide public verifiability, while the other one uses boneh. Both the schemes support block less verification but fail to provide privacy of the DOs data. Block less verification requires linear combination of sampled blocks which gives a clue to TPA to extract the data. To preserve privacy of the data owner supporting block less verification proposed a public auditing scheme and extended that to support batch auditing further. As a result, TPA can simultaneously perform multiple auditing requests from different DUs. But, all these schemes fail to support data dynamics. Moreover, as signatures of the data blocks contain index number of the corresponding blocks, if one block is updated (inserted/modified/deleted), the corresponding verification meta-data (signature) of all other blocks need to be updated. The scheme proposed in uses index hash table (IHT) to support data dynamics in public auditing mechanism reducing the update overhead. Unfortunately, this scheme fails to support batch auditing property. Their previous technique to support data dynamics. Yang et al. proposed an efficient and secure dynamic auditing protocol that achieves all essential features of public auditing.

Also it consumes lesser computation and communication cost. Public auditing scheme for verifying data integrity in the cloud is proposed. Although this scheme does not require certificate for key generation, it fails to achieve privacy, data dynamics, and batch auditing properties. But schemes are based on pairing based cryptography, which requires more verification cost in audit phase.

II. RELATED WORKS

Existing techniques reported in the literature are having high computation and communication costs and are vulnerable to various known attacks, The TPA verification which reduce their importance for applicability in real-world environment. In our existing system implement the data security level at using the Hash Key Function for security key purposed and Encryption method using ECC cryptography Functions. The Each time server change the data keys are changed. For frequent item sets lack a mechanism that enables automatic parallelization, load balancing, data distribution, and fault tolerance on large clusters. Scalability and load balancing challenges in the existing parallel mining algorithms for frequent item sets. Traditional analytical tools that are designed to ingest, visualize, analyze and reproduce results by processing information, are often overwhelmed by the quantity of data collected by domains like Engineering Research, Business, Banking and Finance, IOT, Scientific Research, Health Care etc. With passing time, the quantity of data increases exponentially adding to workload of the platforms Challenges like storing, effective analyzing, managing and procuring quick and correct results from data have to be faced before the analytical tool can extract beneficial information from All the stashed data thereby aiding domains make smarter decisions. These procedures from massive sets require to be performed in a distributed environment. But for establishing clusters like these, a large investment is an obligatory requirement; leaving it out of reach of small and medium enterprises, research etc. Also there is the additional cost of maintenance, space, storage, disaster recovery and so on. Numerous processes have to be checked off before a data analyst is able to procure worthy information from large datasets. Firstly, the data that interests the domains are explored for and a model is built. Multiple iteration of model-building and evaluation may be needed before the analyst is satisfied. This model is tested for processing information and making decision before feedback is attained and Fu rather improvements made on the model until satisfactory performance is achieved.

III. PROPOSED SOLUTION

We propose a protected and effective security saving provable information possession scheme (SEPDP) for cloud storage. It works in three stages, to be specific, key age, signature age and evaluating stage. Most appealing element of SEPDP is that it doesn't utilize any serious calculation like matching based task. Further, we stretch out SEPDP to help numerous information proprietors, group examining, and dynamic information activities. A probabilistic investigation to distinguish the respectability of the squares put away at CSP. We assessed the execution of the proposed plan and contrasted and a portion of the current prominent instruments. We see that the absolute time for check completed by TPA in the proposed plan is not as much as that of the current plans. This implies SEPDP is productive and reasonable to execute the confirmation at the low fueled gadgets. Likewise, the casual security demonstrates that the proposed plan can withstand different possibilities assaults against detached and dynamic enemies. To additionally fortify the security, the proposed plan is reproduced for the formal security confirmation utilizing the popular AVISPA apparatus. The reenactment results guarantee that the plan withstands both replay just as man-in- middle attacks. A thorough similar investigation for the correspondence and calculation costs alongside security and usefulness highlights has been performed among the proposed plan and other existing plans. The relative comparative study shows that the performance of the proposed scheme is better than other schemes.

IV. Architecture Diagram

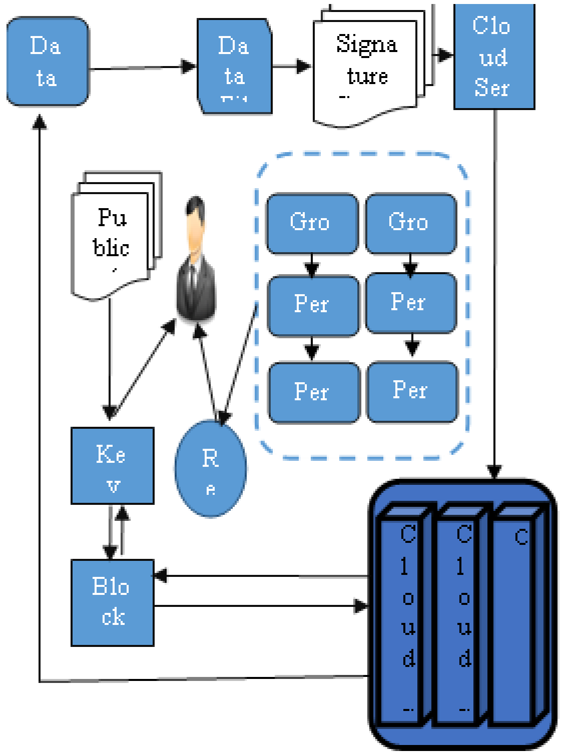

Fig. 1. architecture of SEPDP

V. Methods

A. Data Signature Creation

B. User-Group Creation and Validation.

C. TPA Auditing.

D. Block Rule Generation.

E. Data Encryption.

F. Receiver Response.

G. Data Signature Creation

The Data Owner upload the Datas. This Data was stored in Cloud Storage with Data Tokens and Owner Signatures. The signature was created automatically. The Data owner is fixe the signature.

H. User-Group Creation and Validation.

There lot of cloud user are there the main two users are there Private and Public Cloud users. We implement this two users authorities. We create the groups. In each groups many number of users are there. This users gives the tokens in TPA.

I. TPA Auditing

The TPA Audit the data owner uploaded and downloaded file and file signature.A Third Party Administrator (TPA) is a service organization that provides a variety of services to the insurance industry in accordance with a service agreement. TPAs are usually utilized to provide services associated with employee benefits such as insurance related services to both insurance providers and companies that provide insurance to their employees.TPAs present a huge risk to user organizations (companies using a particular TPA) since TPAs may be processing millions of dollars worth of benefit claims for their clients. User organizations need assurance that the TPAs internal controls are designed and operating effectively to provide the outsourced benefit services.

An independent audit of a TPA is one way to gain assurance regarding the TPAs internal control environment. TPA audits may include detailed tests of claims processed during a particular period of time, data analysis to identify trends and irregularities, and contract analysis. The TPA main goal is monitor the file database, uploaded and downloaded files. And receiver details auditing process. After receive the file is faced on block chain rule and it move on encrypted system The each block files are encrypted and stored on file databases for security purposed. If the client send the request for download the TPA audit the file name signature and username, e mail id and after it permit to download the particular file form database.

J. AES Encryption

The AES is Advanced Encryption Standard key is a Symmetric encryption key this is key is based on ANSCI code formation. This algorithm block key size is 128bit to 256bit. This cryptography is more secured. In our process the cloud data is secure encrypted at using AES encryption. After the encryption process is completed the AES key is generated.

K. Block Rule Generation.

The main purpose of the block rule is security. Once the TPA audit the file. The file is move on block chain rules. In that block chain rule the file is split is size based in each block the each key is generated. The block key is unique key its size is 8 to 16 bit integer values. This block key is monitor by TPA. Once the block key is generated is not changed. The Advanced Encryption Standard, or AES, is a symmetric block cipher chosen by the U.S. government to protect classified information and is implemented in software and hardware throughout the world to encrypt sensitive data.The National Institute of Standards and Technology (NIST) started development of AES in 1997 when it announced the need for a successor algorithm for the Data Encryption Standard (DES), which was starting to become vulnerable to brute-force attacks.

This new, advanced encryption algorithm would be unclassified and had to be ”capable of protecting sensitive government information well into the next century,” according to the NIST announcement of the process for development of an advanced encryption standard algorithm. It was intended to be easy to implement in hardware and software, as well as in restricted environments (for example, in a smart card) and offer good defenses against various attack techniques.

L. Receiver Response.

In this receiver send the file name, signature and password email. This information is received from TPA. The TPA audit this all information and it check authority details if the permission details is granted the TPA send the encrypted key and block chain key. The received has receive this two key the datas was automatically decrypted. At finally the graphical report is shown.

VI. CONCLUSION

The privacy preserving provable data possession scheme (named SEPDP) for untrusted and outsourced storage system is presented. The most appealing features of the proposed scheme is to support all the important features including block less verification, privacy preserving, batch auditing and data dynamics with lesser computation overhead.

REFERENCES

[1] J Monika, M Chui, B Brown, J Begin, R Dobbs, C Rexburg, AH Byers. Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute, 1-137, 2011.

[2] Wei Fan and Albert Bidet. Mining Big Data: Current status, and Forecast to the Future. SIGKDD Explorations 14(2):1-5, 2012.

[3]] Bo Pang and Lillian Lee. Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval. Vol. 2, No 1-2 (2008)1135.

[4] Junky Xuan; Xiangfeng Luo; Jibe Lu, ”Mining Websites Preferences on Web Events in Big Data Environment,” Computational Science and Engineering (CSE), 2013 IEEE 16th International Conference on , vol., no., pp.1043,1050, 3-5 Dec. 2013 doe: 10.1109/CSE.2013.152

[5]] Sanjeev Pippa, Laisha Bart, AK hila Krishna, Hana Gupta, Kuna Arora, Data mining in social networking sites: A social media mining approach to generate effective business strategies, International Journal of Innovations Advancement in Computer Science (IJIACS), Vol. 3, Issue 2, April 2014.

[6] G Nandi and A Das, A survey on using data mining techniques for online network analysis, IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 6, No 2, November 2013.

[7] M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.

[8] Rahul Shoran Reno, Gregory Mock, Ashram Konner, Use of Big Data and Knowledge Discovery to Create Data Backbones for Decision Support Systems, Procardia Computer Science, Volume 20, 2013, Pages 446-453, ISSN 1877-0509 http://dx.doi.org/10.1016/j.procs.2013.09.301.

[9] Prekopcsk, Sultan, et al. ”Radon: Analyzing big data with rapid miner and Hadoop.” Proceedings of the 2nd Rapid Miner Community Meeting and Conference (RCOMM 2011). 2011

[10] Jaliyah Ekanayake, Hue Li, Binging Zhang, Thaliana Gunarathne, Seung-Hee Bae, Judy Qi, and Geoffrey Fox. 2010. Twister: a runtime for iterative Mapreduce. In Proceedings of the 19th ACM International Symposium on High Performance Distributed Computing (HPDC ’10).

[11] Kong-Ha Lee, Yoon-Jon Lee, Hyunsik Choi, Yon Don Chung, and Bangka Moon. 2012. Parallel data processing with Mapreduce: a survey. SIGMOD Rec. 40, 4 (January 2012),120.DOI=10.1145/2094114.2094118 http://doi.acm.org/10.1145/2094114.2094118

[12] Shvachko, K.; Haring Kiang; Radial, S.; Chandler, R., ”The Hadoop Distributed File System,” Mass Storage Systems and Technologies(MSST), 2010 IEEE 26th Symposium on , vol., no., pp.1,10, 3-7 May 2010 doe: 10.1109/MSST.2010.5496972

[13] Chantal Yadav, Shulman Wang, Manor Kumar, Algorithm and approaches to handle large Data- A Survey, International Journal of computer science and network, vole 2, issue 3, 2013

[14] A K Jain, M N Marty, P. J. Flynn, Data Clustering : A Review, ACM COMPUTING SURVEYS, 1999

[15] B.A Tide, R.G Mehta, D.P Rena, A Novel Approach for High Dimensional Data Clustering in International Journal of Engineering Science and Advanced Technology (IJESAT) ISSN 22503676 Vol.02(3) May-Jun 2012

[16] Tina Zhang, Raghu Ramakrishna, and Myron Livy. 1996. BIRCH: an efficient data clustering method for very large databases. SIGMOD Rec. 25, 2 (June 1996), 103-114. DOI=10.1145/235968.233324 http://doi.acm.org/10.1145/235968.233324