Cognitive highly distributed Cooperative Provable Data Possession scheme for achieving Dynamic Data Operations in Multi-Cloud Storage

Tenzin Chozom
Vels University, Chennai

Dr. Mayilvahanan
Vels University, Chennai

Dr. A.Muthukumaravel
Vels University, Chennai

Abstract - Cloud computing denotes to applications and services that run on a distributed network along with all the associated standards and protocols that provide a set of services to the clients. Provable Data Possession (PDP) is an existing scheme for ensuring the possession of outsourced data on single cloud storages; which doesn't support the dynamic scalability. In this paper, it address the construction of an efficient PDP, which holds the scalability of service through the hybrid cloud service providers to cooperatively store and maintain clients' data. This paper proposes Cooperative PDP (CPDP) model that comprises three effective mechanisms i) Multi-cloud storage: This mechanism allows to store the large size files distributively among the multiple clouds by dividing into number of blocks that achieves scalability ii) Hash Index Hierarchy: it provides a standard representation for the divided blocks for file storage and also derives the relationship between the blocks to improve data accessibility. iii) Homomorphic Verifiable Response: it is a challenge response protocol used to integrate multiple responses from the different cloud service providers that facilitates to effectively locate the outsourced data in distributed multi-cloud storage and also reduces the communication cost and storage overhead. In addition to, the major contribution of this paper is achieving dynamic data operations on the outsourced data with high security and flexibility. Experimental setup is carried out based on java prototype implementation and also evaluate the effectiveness of proposed approach in terms of communication cost and integrity verification time.

Keywords: Data Security, Provable Data Possession, Multi-Cloud Storage, Cooperative PDP, Dynamic Data Operation

I. INTRODUCTION

In recent years, cloud computing is becoming popular and crucial in each and every ones' life. Indeed, cloud computing is something that have been utilizing for a long time; it is the network based computing in which takes the technology, services and applications that are similar to those on the Internet and turns them into self-service utility. Apart from the effectiveness of cloud computing, one of the core design principle is that which deals its performance in a peak that is dynamic scalability, which ensures cloud storage service to handle large amount of data in a flexible manner and to be readily enlarged by integrating the public and private clouds. In this paper, a cloud provider takes the concept of distributed cloud storage environment represents a family of multi-cloud (or hybrid cloud) for handling the clients' outsourced data. Since the outsourced data could access by all the clients other than authenticate client from the cloud storage provider thus it would bring irretrievable losses to the clients. Thus, it is very important for cloud service providers to bring out the security and privacy protection techniques on outsourced data.

Ateniese et al.[2] Provable Data Possession (PDP) and Proofs Of Retrievability (POR)[4] is a guarantee technique for a cloud service provider to ensure the possession of outsourced data. This technique is not suitable for large size of files. The enhanced PDP scheme have been developed, such as Scalable PDP[3] and Dynamic (PDP)[5] , its

continuing on same PDP effects at untrusted storage in a single cloud storage provider and it doesn't support for a hybrid cloud environment. The existing PDP scheme has been developed within single clouds offers publicly verifiable version, which allows anyone, not only the owner to challenge the untrusted server for data possession. However, these schemes are vulnerable to security attacks cause of the depending on numerical scale of blocks.
In this paper, it supports a hybrid cloud environment comprised of public and private clouds that outsourced file system share some same features: a single metadata service provider furnishes centralized management by a global namespace: files are divide into blocks and stored on a block servers; and the system are comprised of interconnected cluster of block servers. Those features allow cloud service providers to maintain dynamic scalability and data accessibility. Through Cooperative Provable Data Possession (CPDP) scheme, dynamic data operation such as Insertion, Deletion and Update can be achieved after verifying the integrity of outsourced data by an authenticated client.

1.1 Aim & Objective:

The main aim and objective of this paper (i) to construct an efficient CPDP scheme for distributed multi-cloud storage that achieves high dynamic scalability of service and data availability.(ii)To cooperatively store and maintain the clients' outsourced data over the multi-clouds with high degree of privacy and security. (iii)To present an effective integrity verification framework that achieves privacy protection on the basis of best cryptographic scheme. (iv)To ensure the security in CPDP scheme against data leakage attack and tag forgery attack.

1.2 Contribution:

This paper presents a highly distributed multi-cloud storage to achieve a dynamic scalability of service; which is unsatisfied in existing Provable Data Possession (PDP) scheme. To meet this goal, first propose a verification model for hybrid cloud storage along with two effective mechanisms: Hash Index Hierarchy (HIH) and Homomorphic Verifiable Response (HVR) that achieves dynamic scalability of service and privacy protection on outsourced data. It is highly suitable for distributed cloud storage system. Moreover, the homomorphic verifiable response incurs limited communication cost and operation to perform integrity verification using multiple challenge response protocol.
To improve the system performance with respect to proposed scheme, dynamic data operation such as Insertion, Deletion and alteration to be achieve through Cooperative PDP (CPDP) technique. In CPDP, once the integrity verification is done by authenticate client then the respective authenticate client could modify their own verified outsourced data by applying any data operation such as insert, delete or update on their own data stored on a distributed storage server along with the key and signature function. Finally, this proposed scheme could achieve high scalability, high security and better performance.

1.3 Paper Organization:

In the remaining paper is organized as follows. In section 2, providing the detailed explanation of proposed scheme. In section 3,describing the methodology utilized in this paper. In section 4, describing the performance evaluation of this scheme. In section 5, explaining the underlying techniques, which are used in the construction of this scheme. In section 6, provides the conclusion of this paper.

II. OVERVIEW OF PROPOSED COGNITIVE HIGHLY DISTRIBUTED COOPERATIVE PROVABLE DATA POSSESSION SCHEME:

2.1 DEFINITION AND FRAMEWORK FOR MULTI-CLOUD:

To achieve high scalable, high security, low cost and high performance, this ensures the cloud storage service to handle large files in a flexible manner. Although existing PDP schemes offers publicly accessible version to any user, not just the owner to access the data from cloud server thus server can deceive the owners and it is insecure against attacks and moreover it do not fit for multi-cloud storage. In order to address this existing problem, this paper presents a cooperative provable data possession (CPDP) which supports the hybrid cloud environment to store and maintain the clients' data. The hybrid cloud storage service involving three different entities. The cloud client, who stores or access the data in the cloud; the cloud service providers(CSP) , which has enough memory space and computation resources to manage and provide storage services; the trusted third party(TTP), who stores the clients' inspect data and offers the query services for their data.

The two effective mechanisms for constructing this CPDP scheme: hash index hierarchy (HIH) which provides hierarchical representation of stored data that leads to achieve high data accessibility and homomorphic verifiable response (HVR) used to effectively locate the outsourced data among the distributed cloud storage using challenge response protocol. Due to the deployment of above two techniques in this approach aids to provide dynamic scalability and flexibility. It is highly secure, transparent verification and high performance than non-cooperative approaches. In addition to CPDP scheme, it's also possible to achieve dynamic data operation such as Insertion, Deletion and Alteration on outsourcing data by an authenticate client. Dynamic Data operation allows clients to easily modify their own existing outsource data on hybrid cloud storage.

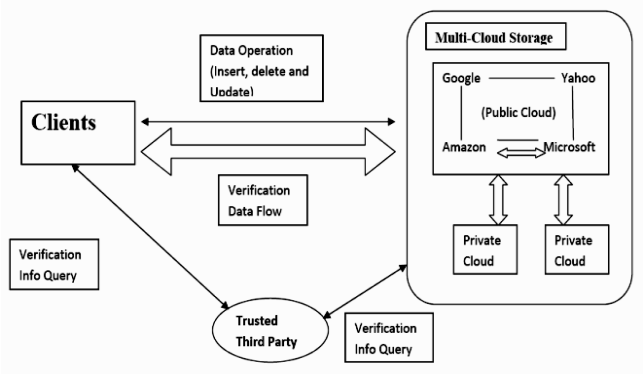

Figure 1: An overview of Proposed CPDP scheme

III. METHODOLOGY:

In this section, this paper presents the following technique: multi cloud storage, Hash Index Hierarchy (HIH), Trusted Third Party (TTP), Cooperative PDP, Homomorphic Verifiable Response (HVR) and Dynamic Data Operation.

3.1 Multi-Cloud Storage:

As existing PDP scheme offers publicly accessible version and it is vulnerable to attacks such as data leakage attack and tag forgery attack. It has single cloud storage service to store and maintain the clients' data therefore existing PDP scheme is not suitable for multi-cloud storage. To address this problem, the proposed scheme (CPDP), provides a multi-cloud storage service and a client's data are divided and stored in multi-cloud storage (Public and Private clouds) with unique key and signature to achieve high data availability and data accessibility. Moreover, as cooperative PDP supports multiple CSPs(cloud service provider) to cooperatively stores and maintain the clients' data and it is used to ensure the possession and accessibility of data stored in all the CSPs.

3.2 Hash Index Hierarchy (HIH):

While storing outsourced data to distributed multi-cloud service providers, the outsourced data represents hierarchical structure, which has three layers and derives the relationship among all the blocks to improve data accessibility. The following three layers are: Express Layer, which provides naming representation of stored resources, Service Layer dealing with the multi-cloud storage services. And Storage Layer which understand the data storage on multi-cloud server.

In this paper, in express layer the data are divided and stored into distributed multi-cloud service provider and the each multi-cloud service provider is shown uniquely in service layer. And in storage layer, each multi-cloud service provider fragments and stores the attributed data into multi-cloud storage server. It achieves the verification of data integrity for outsourced storage. In this layer, an outsourced file F is divided into m-blocks (N1,N2……Nm) and each block Ni is split into sectors Se and the each sector Se represents to tags thus the increasing of sectors Se will be effecting the storage of signature tags. And randomly verify the correctness of file which highly suitable for large files. The resource in Express Layer are split and stored into various CSPs in Service Layer .And each distributed multi-cloud provider fragments and stores the allotted data into the storage server in Storage Layer.

3.3 Trusted Third Party (TTP):

It is a kind of server, which communicates both with the client and cloud service provider to achieve the security on the multi-cloud environment. It stores a set of public verification information. And it is trusted to stores verification parameters and offer public query services for clients' parameters. TTP server is developed as a trust base on the cloud to achieve the high security and performance.

3.4 Cooperative PDP:

The possession of data stored in multi-cloud environment can be achieved through CPDP based on interactive proof system (IPS). A cooperative provable data possession P=(KeyGen, TagGen, Prf) is a accumulation of two algorithms (i.e. keyGen,TagGen) and an interactive proof system Prf. When the client provides the Key and Tag as input to the TTP (Trusted Third Party) and Trusted Third Party ensures the integrity of data stored on cloud servers along with the Tag matches to its own outsourced data.

1)KeyGen(1N): takes a unique random nounce N as a input and returns a private key and public key- pair(Private, Public)

2)TagGen(Private,Fi,S): takes input as a private key Private , a file Fi and a group of cloud service provider S={SN}, and returns the triplets (Ψ,δ,λ), where Ψ is the unknown in tags, δ=(v, U) is a group of verification arguments v and an hierarchy U for Fi, λ={λ(N)} SN $\in$S refers a set of all tags ,λ(N) is a tag of the divide D(N) of Fi in SN.

3)Prf(S,V): is a proof of data possession between various distributed multi-cloud service provider(S={SN}) and a verifier((V), i.e.,

(Σ (SN$\in$S) SN(D(N), λ(N))$\leftrightarrow$ V)(Public,δ)

0 D={D(N)} is integral

1 D={D(N)} is varied

As each SN takes input as a file Fi(N) and a set of tags λ(N) and a public key Public and a set of public arguments δ are the common input between S and V. At the last of protocol operation, V returns a bit {0|1} referring FALSE and TRUE, Where Σ(SN$\in$S) denotes cooperative computing in SN$\in$S.

By sequentially, to check the data stored in each distributed multi-cloud in cooperative provable data possession i.e.

Λ(SN$\in$S)$\langle$SN(D(N), λ(N))$\leftrightarrow$V$\rangle$(Public,δ)

"Λ" refers to the logical AND operations between all the Boolean results of $\langle$Public,$\delta\rangle$ $\forall$ SN$\in$S. Finally it would to achieve communication and storage overheads

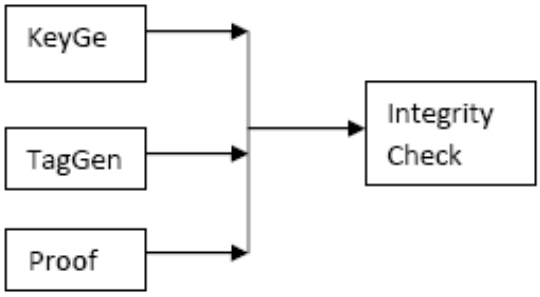

Figure 2: Cooperative PDP model

3.5Homomorphic Verifiable Response(HVR):

It is a one of the most important mechanism of CPDP to conceal the location of outsourced data in distributed cloud storage environment. And the main core principle is to reduce the communication bandwidth. It is used to integrate multiple responses from the different CSPs in CPDP scheme.The multiple challenges from CSPs taken as messages and combined to form a unique response with less communication overheads. Its homomorphic properties to whole data and tags to a fixed size response to minimize the network communication overhead.

It is a challenge response protocol used to integrate multiple responses from the different cloud service providers that facilitates to effectively locate the outsourced data in distributed multi cloud storage.

3.6 Data Operations:

The proposed scheme can efficiently manage completely dynamic data operations i.e. Data Alteration (U), Data Deletion (De) and Data Insertion(Is) for distributed multi-cloud storage. They are:

Data Alteration: It refers to the replacement of existing blocks with new ones. It is highly often used to achieve data correctness and enhancement on a distributed multi-cloud storage. Imagine that the client needs to update the existing file in(i.e m-th block) bm to bm′ . Based on the new data i.e. bm′, the user (client) generates the signature and update message with the arguments U, m, bm′, signature and sends to the multi-cloud storage server where U refers the alteration operation. And the server executes the update message and serve replaces the block bm to bm′. Finally, clients requires to provide proof for authentication of existing block to the multi- cloud storage server.

Data Deletion: It is widely used operations when the error occurs on existing file in distributed multi-cloud storage. It refers to deleting the existing file which acts as an error or fault. To delete the single block, then specified block will be deleted and proceeds the next blocks one block forward. And once the server receives the update message for deleting block bm then bm will be deleted from its memory space.

Data Insertion: It refers to inserting new blocks in some specified positions in file F. When the client wants to insert new block bT in the m-th block bm. Based on new block bT, the client generates signature and constructs a update message with insertion operation and sends to the distributed multi-cloud storage server. Then server inserts the new block bT and finally clients needs to provide the proof for this operation

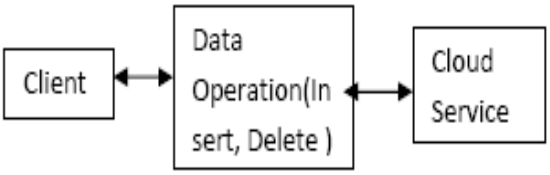

Figure 3: Data operation achieved through CPDP model

IV. PERFORMANCE ANALYSIS:

To formalize the effects of the proposed scheme in comparison with existing scheme (PDP), the proposed scheme is highly based better performance with high security. In this section, presents an experimental result in comparison with non-cooperative approach. It achieved the dynamic data operation in distributed multi-cloud data storage to support better performance and high secure. This experiment is carried out by using Java jdk 1.6 on a system with an Intel(R) Pentium(R) Dual core processor operating at 2.00GHz, 2GB RAM running windows 7 and the full java codes are tested on window 7 platform.

This paper presents two graphs representation of experimental results of proposed scheme in comparison with Provable Data Possession (an existing scheme) in distributed multi-cloud data storage. In Figure 4, it conveys that the existing scheme (PDP) carries large communication and computation cost as the file size increases where as the proposed scheme (CPDP) takes lesser communication and computation cost.

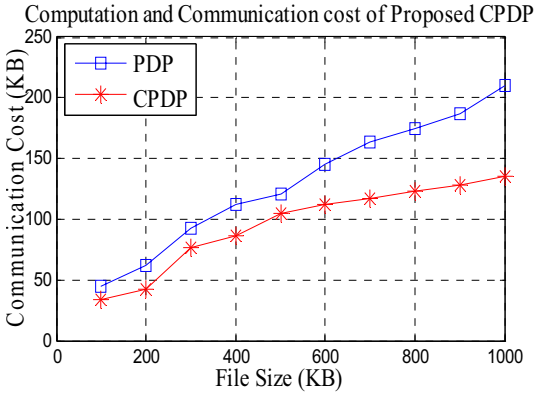

Figure 4: Relationship between computational cost and the number of sectors in each block.

This paper comprises very effective mechanism i.e. Homomorphic Verifiable Response(HVR), it is a challenge response protocol used to integrate multiple responses from the different cloud service providers that facilitates to effectively locate the outsourced data in distributed multi cloud storage and also reduces the communication, computation cost and storage overhead.

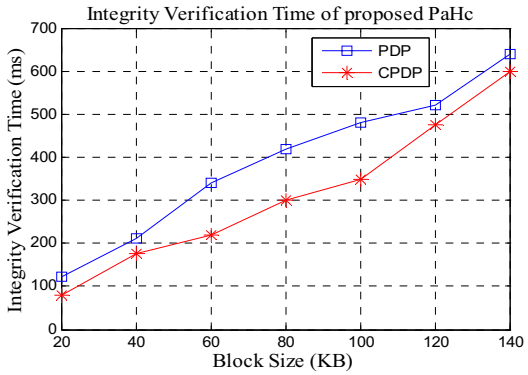

Figure 5: Comparison of integrity verification time between PDP and CPDP

In this above Figure 5, the proposed scheme (CPDP), takes less integrity verification time than non-cooperative approach (i.e. PDP). As the proposed scheme based on one of the effective mechanism i.e. Hash Index Hierarchy, provides a standard representation for the divided blocks for file storage and also derives the relationship between the blocks to improve data accessibility.

V. RELATED WORKS:

The two most popular basic approaches called Provable Data Possession (PDP) [2] and Proofs of Retrievability (POR) [4]. Ateniese et al. [2] proposed the PDP scheme to prove the integrity and authenticity of data within a single cloud storage provider. They offered a publicly verifiable version that anyone can access the outsourced data not only the owner from the single cloud storage provider for data integration and possession. It is highly vulnerable to security attacks and moreover it imposes a significant I/O and computational burden on the server. And it is not suitable for hybrid cloud environment.

Steven Y al. [6] presented a availability of intermediate data in cloud computation, in which the intermediate data that is generated during dataflow computations within clouds was managed efficiently and the intermediate storage system (ISS) provides the requirements for the design phase. Unfortunately, it's difficult to provide multiple stages of computation and to provide connection for communication. Shacham and Waters[7] proposed an enhanced version called Compact POR, which gives the first proof of retrievability schemes with full of proofs of security against arbitrary adversaries in the strongest model but have to depend on other function to get the attributes and also have complexity in computational and communicational. It could not implement against leakage of data blocks in the verification process.

Bowers et al. developed a HAIL (High-Availability and Integrity Layer) [8], which is a distributed cryptographic system that PDP problem can be easily solved by set of servers and it's based on a integrity protected error correcting code (IP_ECC) and HAIL manages file redundancy across cloud storage providers hence this system is more suitable for RAID(is mainly for crash recovery) instead of clouds storage. To achieve dynamic data operation, Ateniese et al. proposed a Scalable PDP [3], which constructing a highly efficient and provable secure PDP technique based entirely on symmetric key encryption, but the cloud storage server can be dishonest with owners and user cannot perform any operation.

VI. CONCLUSION

To address the existing PDP problem of integrity verification in a single cloud storage. And this paper constructs a cooperative Provable Data Possession (CPDP) which supports distributed cloud storage in a hybrid cloud environment and it achieved the dynamic scalability and data accessibility.

This proposed scheme, which offered the dynamic scalability by integrating the public and private clouds along with two effective mechanism i.e. Hash Index Hierarchy (HIH) and Homomorphic Verifiable Response(HVR). And it improved the data accessibility by representing hierarchical structure for the divided blocks and maintains the relationship among them. It introduced a less communication, computational and storage overheads and finally, a

cognitive highly distributed cooperative provable data possession scheme achieved dynamic data operation on the outsourced data with high security. With that high security and better performance analysis depict the proposed scheme is highly secure and flexibility.

REFERENCES

[1] M.Armbrust,A.Fox,R.Griffith,A.D Joseph, R.H Katz, A.Konwinski,G.Lee, D.A Patterson, A.Rabkin, I.Stoica, and M.Zaharia"Above the clouds: A Berkeley view of cloud computing," EECS Department, University of California, Berkeley, Tech.Rep.,Feb 2009.

[2] G.Ateniese, R.C. Burns, R.Curtmola, J.Herring,L.Kissner, Z.N.J. Peterson, and D.X.Song, "Provable data possession at untrusted stores," in ACM conference on computer and communications security, ACM, 2007.

[3] G.Ateniese,R.D. Pietro, L.V. Mancini, and G. Tsudik, " Scalable and efficient provable data possession" in proceedings of the 4th international conference on security and privacy in communication networks, SecureComm, 2008.

[4] A.Juels and B.S.K. Jr.,"Pors: Proffs of retrievabilityfor large files,"in ACM conference on computer and communication security, ACM, 2007.

[5] C.C.Erway, A. Kupcu, C. Papamanthou, and R.Tamassia, "Dynamic provable data possession," in ACM conference on computer and communication security, ACM, 2009.

[6] Steven Y.Ko, ImranulHoque, Brian Cho, Indranil Gupta, "On Availability of Intermediate Data in cloud computations"

[7] H.Shacham and B.Waters, "Compact Proofs of retrievability" in ASIACRYPT, ser.Lecture notes in computer Science, springer, 2008

[8] K.D .Bowers,A.Juels, and A. Oprea, "Hail: a high availability and integrity layer for cloud storage,"in ACM conference on computer and communication security, ACM, 2009.

[9] Y.Zhu,H.Wang,Z.Hu,G.J.Ahn,H.Hu, and S.S.Yau,"Dynamic audit service for integrity verification of outsourced storage in clouds"ACM, 2011

[10] B. Sotomayor, R. S. Montero, I. M. Llorente, and I. T. Foster, "Virtual infrastructure management in private and hybrid clouds," IEEE Internet Computing, vol. 13, no. 5, pp. 14–22, 2009.

[11] D. Boneh and M. Franklin, "Identity-based encryption from the weil pairing," in Advances in Cryptology (CRYPTO'2001), vol. 2139 of LNCS, 2001, pp. 213–229.

[12] Ronald L.Krutz and Russell Dean Vines, "Cloud Security- A Comprehensive Guide to secure Cloud Computing", copyright @2010 by Wiley India Pvt.Ltd., 4435-36/7, New Delhi-110002.

[13] William Stallings, "Cryptography and Network Security", PHI, 2006

[14] M Steen Strub, "Routing in Communication Networks", PH International, NY 1995.

[15] Behrouz A Ferouzan, "Data Communications and Networking (3/e)" TMH, 2004

[16] Charlie Kaufman, Radia Rerlman Mike Specines, "Network Security – Private Communication in a Public World", PHI (2/e) 2002.