Improving Data Spillage in Multi-Cloud Capacity Administration

[1]Nivedha.D, [2]Vinitha.R

[3] Mrs. N.Gowri Vidhya, M.E Associate professor,

Department of Computer Science and Engineering,

Prince Dr K Vasudevan college of Engineering And Technology, Chennai, India

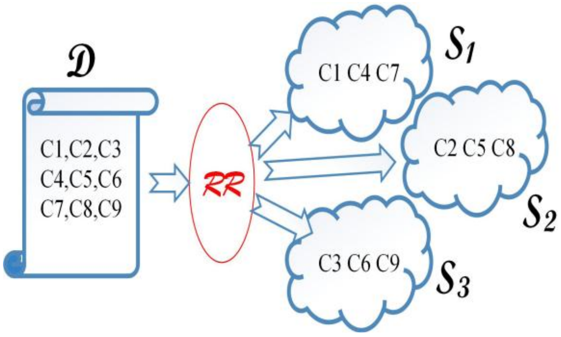
---***---

Abstract--In provable multi-cloud, is used to prevent the data from the data leakage. First they should register their details in the user registration and in owner details. If user want to upload a file, it may be any kind of file. Here we are using AES/ABS algorithm to upload a file. Then it will be compressed by using compressor. After compressed give the new file name and then it will be encrypted and data will be splitted by using Fully Homomorphic Encryption(FHE). The splitted files will be stored in different database. After finishing the process it will automatically generate two secret key. If they user want to download the file or view the file, by giving the new filename and the two- secret key. StoreSim, an information leakage aware storage system in multicloud. StoreSim aims to store syntactically similar data on the same cloud, thus minimizing the user's information leakage across multiple clouds. It reduce access timeand communication from users. Convert zip while upload the data. The authorized users send the file request to cloud server again server send the encrypted data to authorized user. And authorized user get the decrypt key from data owner. we introduced new technique that is Fully Homomorphic Encryption(FHE) for take Multi copy of data, File security, Data Corrupted. In this Project we have to FHE algorithm for protect the data. That are keygen, copygen and taggen. Above the process done in Existing system using Single copy of Dynamic Data. StoreSim, an information leakage aware storage system in multicloud. StoreSim aims to store syntactically similar data on the same cloud, thus minimizing the user's information leakage across multiple clouds.

Keywords—Cloud computing , Searchable encryption , privacy- preserving , Keyword search , Ranked search.

I. INTRODUCTION

Cloud computing is a new form of internet based computing resources and data to computes and other devices on-demand. It is the hostel services over the internet. Services can be public, private and hybrid services. It is used to provide a service paradigm of user workloads. The three main purpose of cloud computing are self-services provisioning, elasticity and pay per use .The cloud computing infrastructure as a service, platform as a service and software as a service. This architecture consists of front-end platform called cloud clients. It is a model for enabling shared pool of configurable computing resources. Security and privacy is the biggest concern about cloud computing. Since data management and infrastructure management in cloud is provided by third-party, it is always a risk to handover the sensitive information to such providers. Although the cloud computing vendors ensure more secure password protected accounts, any sign of security breach would result in loss of clients and businesses. LOCK-IN is very difficult for the customers to switch from one Cloud Service Provider (CSP) to another. It results in dependency on a particular Cloud Service Provider for service.



Security and privacy is the biggest concern about cloud computing. Since data management and infrastructure management in cloud is provided by third-party, it is always a risk to handover the sensitive information to such providers. Although the cloud computing vendors ensure more secure password protected accounts, any sign of security breach would result in loss of clients and businesses. LOCK-IN is very difficult for the customers to switch from one Cloud Service Provider (CSP) to another. It results in dependency on a particular Cloud Service Provider for service. Isolation failure risk involves the isolation mechanism that separates storage, memory, routing between the different tenants.

In deployment model, public cloud performs the service providers, provides services for all the users or organization whenever is making use of it and easy to implement. Private cloud, like the name suggest, owned by particular institution. Hybrid cloud ,contains both public and private cloud and more over it can ensure safety, scalability and performances.

Isolation failure risk involves the isolation mechanism that separates storage, memory, routing between the different tenants. Manipulate and configure the application online at any time. It does not require to install a specific piece of software to access or manipulate cloud application. Cloud Computing offers online development and deployment tools, programming run-time environment through Platform as a Service model. Cloud resources are available over the network in a manner that provides platform independent access to any type of clients. Cloud Computing offers on-demand self-service.

The resources can be used without interaction with cloud service provider. Cloud Computing is highly cost effective because it operates at higher efficiency with greater utilization. It just requires an Internet connection. Cloud Computing offers load balancing that makes it more reliable. Risks Although Cloud Computing is a great innovation in the world of computing, there also exist downsides of cloud computing.

II. PROPOSED SYSTEM

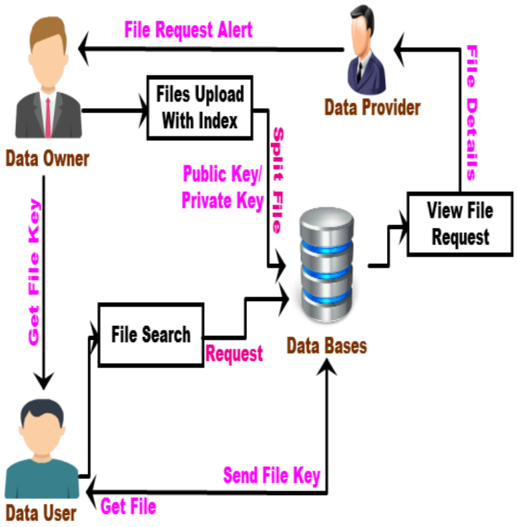
In Provable Multi files Dynamic data Possession in cloud computing deals with stored data in Dynamic way to cloud server. Multi files Means, data to be copied in multiple server. In this the owner to upload the data in cloud server with automatically data to take multiple files then that files are stored in multiple server. If upload the data in multi server to avoid the data loss from Hacking and server crash.

The uploaded data are stored in multiple server (Multi files).In proposed one scheme FHE algorithms are used. If Owner upload the data, automatically prepare three files then stored in three server that why for security and avoid server overload. That copies also encrypted so cloud service provider or any others cant hack the data. If owners upload the data, server automatically convert to zip formats. So server reduce the file size automatically. Owner share the file to authorized user. Then authorized user send the file request to cloud server again server send the encrypted data to authorized user. And authorized user get the decrypt key from data owner.

Introduced new technique that is Fully Homomorphic Encryption(FHE) for take Multi files of data, File security, Data Corrupted. In this Project we have to FHE algorithm for protect the data. That are keygen, copygen and taggen. Multi files Data, So reduce access time and communication from users. If one copy corrupted it will be redirect to another server can download the file. Convert Zip while upload the data.

III. SYSTEM ARCHITECTURE

A system architecture is the conceptual design that defines the structure and the behavior of a system. An architecture description is a formal description of a system, organized in a way that supports reasoning about the structural properties of the system.



IV. EFFICIENT MEASUREMENT OF PAIRWISE INFORMATION LEAKAGE

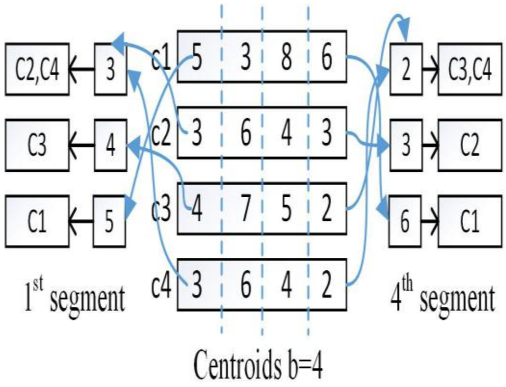
The pairwise information leakage as delta Jaccard similarity. For each pair of data nodes (chunks), we convert the data nodes as sets of words and compute the Jaccard similarity. However, the set operations for measuring pairwise similarity can be quit expensive [10], even assuming small-sized chunks, given that the number of pairs increases quadratically as the number of chunks increases. Thus, we need an efficient algorithm to compute the Jaccard similarity with less computation and storage overhead. In the following, we first introduce the background of MinHash algorithm, which provides a fast way to compute Jaccard similarity, and explain why we cannot apply the existing approaches directly. Next we present BFSMinHash, a Bloom filter sketch for MinHash in order to reduce storage overhead.

A. MinHash Background MinHash [10, 11] uses hashing to quickly estimate the Jaccard similarity of two sets which can be also interpreted as "the probability that a random element from the union of two sets is also in their intersection", P rob[min(h(t1)) = min(h(t2))] = | t1 ∩ t2 t1 ∪ t2 | = P(t1, t2) where h is the independent hash function and min(h(S1)) gives the minimum value of h(x), x ∈ t1. Therefore, we can choose a sequence of hash functions s1, s2, · · · , sk and compute the minimum values of each hash function as MinHash signatures, i.e., Sig(S) = {min(hi(t))|i = 1, · · · , k}. It follows that Jaccard similarity of two sets is approximated as |Sig(t1) ∩ Sig(t2)|/k. However, MinHash with many hash functions needs to compute the results of multiple hash functions for every member of every set, which is computationally expensive. Although it reduces the storage cost greatly, it can still be heavy given the huge number of data nodes.

To reduce the storage overhead reduce the storage space. However, this approach does not work for the MinHash with a single hash function since all the signatures are computed by the same hash function. Instead, we design BFS MinHash, a Bloom-filter sketching scheme for Minhash, which uses a single hash function. MinHash Background MinHash [10, 11] uses hashing to quickly estimate the Jaccard.

similarity of two sets which can be also interpreted as "the probability that a random element from the union of two sets is also in their intersection", P rob [min(h(t1)) = min(h(t2))] = | t1 ∩ t2 t1 ∪ t2 | = J(t1, t2) where h is the independent hash function and min(h(S1)) gives the minimum value of h(x), x ∈ S1. There are three steps in BFS MinHash: shingling, fingerprinting and sketching. Firstly, we convert each data chunk to a set of shingles which are contiguous subsequences of tokens. The process of shingling is to tokenize the byte stream into a set of shingles. For example, if the input is "abcde" and the size of a shingle is 2, the set of shingles is {ab, bc, cd, de}. From this perspective, we only consider the similarity in a syntactic 11.7%.

In our case, we aim to keep the size of Bloom filter as small as possible and therefore the Bloom filter in our BFS MinHash algorithm always employs a single hash function.



V. SECURE INNER PRODUCT COMPUTATION

In KNN scheme [27], the distance between the data record and the query vector. The secret key is composed of (n+1)-bit vector as N and two (n+1) * (n+1) matrices as {$S_{1,}S_2$}, where n is the number of field for each record. A random number R > 0 as (r_q, R). Then $p_i^{->}$ is split into two random vectors ($p_{i,}$ p_i^n) and q is split in to two random vectors (q, q^n), namely if the jth bits of n is 0, p_i^1 [j] and p_i^n [j] are set as same as p_i [j], while q^1[j] and q^n [j] are the set of two random numbers. So that the sum is equal to q[j]; the split data vector pair {p_i^1,p_i^n} is encrypted as {S_1^v,p_i^1,S_2^v,p_i^n}.

The split query vector pair {q^1,q^n} is encrypted as {S_1^v,q_i^1,S_2^v,q_i^n}. The query step, the product of data vector pair and query vector pair, i.e., 0:5R(||piR^2||-2pi.q), is serving as the indicator of euclidean distance (RpiR2 - 2pi. Q+Rq2) to select k nearest neighbors. As the MRSE is using the inner product similarity instead of the euclidean distance, we need to do some modifications on the data structure to fit the MRSE framework. One way to do that is by eliminating the dimension extension, the final result changes to be the inner product as Rpi.q. While the encryption of either data record or query vector involves two multiplications of a n * n matrix and a d-dimension vector with complexity O (n^2) the final inner product computation involves two multiplications of two d-dimension vectors with complexity O (n) . In the known cipher text model, the splitting vector N is unknown so p_i^1 and p_i^n are consider as two random n-dimensional vectors. To solve the data vectors or { S_1 , S_2 }, similarly q^1 and q^n are also consider as two random n- dimensional vectors.

TABLE 1

K_3 Appears in Every Document

Doc	Query {k1,k2,k3}	Query {k1,k2}
1	x1=3,y1=R(3+€$_1$)+t	x_1'=2,y_1'=R'(2+€$_1$)+t'
2	X2=2,y2=R(2+€$_2$)+t	X_2'=1,y_2'=R'(2+€$_2$)+t'
3	X3=1,y3=R(1+€$_3$)+t	X_3'=0,y_3'=R'(2+€$_3$)+t'

Query ($V_{w,k,i}$) with the trap door V_w, the cloud server computes the similarity spose of each document E_i as in (1) WLOG assume that R > 0 . After sorting the scores server becomes top-k ranked in the list E_w.

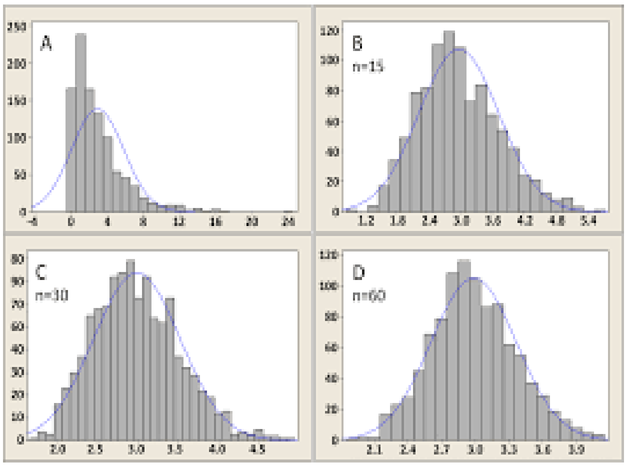
$$J_i \,.\, V_w = \{ S_1^v C_i^1 , S_2^v C_i^n \} . \{ S_1^v D_i^1 , S_2^v D_i^n \}$$

$$= C_i^1 . D_i^1 + C_i^n . D_i^n$$

$$= C_i . D_i$$

$$= (C_i \text{€}_i , 1) . (RD , R, t)$$

$$= R (C_i . D + \text{€}_I) + t.$$



VI. MULTI-CLOUD STORAGE SERVICES

Privacy-Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data. In this to Solve the challenging problem of privacy-preserving multi- keyword ranked search over encrypted data in cloud computing (MRSE). Related works on searchable encryption focus on Single Keyword search (or) Boolean Keyword search. Among various multi keyword semantics, we choose efficient similarity measure of "coordinate matching". Verifiable attribute-based keyword search over outsourced encrypted data. The cloud cannot be fully trusted, so that outsourced data should be encrypted.

To overcome the proposed a novel cryptographic solution, called verifiable attribute-based keyword search (VABKS). The solution allows a data user, whose credentials satisfy a data owner's access control policy, to search over the data owner's outsourced encrypted data, outsource the tedious search operations to the cloud, verify whether the cloud has faithfully executed the search operations. SPANStore which is for cost effective and to minimize an application providers cost we combine three principles. SPANStore spans multiple cloud providers to increase the geographical density of data centers. It satisfies latency goals with the higher storage. .Data propogation cost in order to satisfy fault tolerance and consistency requirements. To implement a task using two phase locking and data propogation.

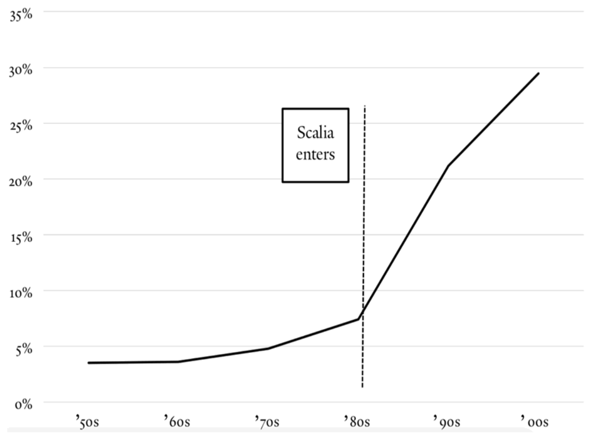


Fig : Scalia versus a fixed set of providers during active repair

SCALIA, a cloud storage brokage solution that continuously adapt the placement of data. To avoid vendor lock-in and to increase availability and durability. Orchestration of a non static set of public cloud and corporate-owned private storage resources. Orchestration which of data can be stored in a sequence manner. Searchable encryption is a cryptographic primitive allowing for private keyword based search. To bridge this gap, we are motivated to propose a practical multi-user searchable encryption scheme. In the approaches we improve the search efficiency at the cost of large storage. A Secure and Dynamic Multi-Keyword Ranked Search Scheme over Encrypted Cloud Data and it supports dynamic update operations like deletion and insertion of documents. Based on special tree-based index structure and propose a "Greedy Depth-first Search" algorithm to provide efficient multi-keyword ranked search.

The secure kNN algorithm is utilized to encrypt the index and query vectors, and meanwhile ensure accurate relevance score calculation between encrypted index and query vectors. In order to resist statistical attacks, phantom terms are added to the index vector for blinding search results. Search over encrypted data is a critically important enabling technique in cloud computing. To protecting user data privacy in the untrusted cloud server environment by using attribute-based encryption (ABE). Our scheme allows multiple owners to encrypt and outsource their data to the cloud server independently.

Fine-grained search authorization is also implemented by the owner-enforced access policy on the index of each file.Users can generate their own search capabilities without relying on an always online trusted authority. Enabling Public Auditability and Data Dynamics for Storage Security in Cloud Computing for to support for data dynamics via the most general forms of data operation, such as block modification, insertion, and deletion, is also a significant step toward practically.

Computing are not limited to archive or backup data only. While prior works on ensuring remote data integrity often lacks the support of either public auditability or dynamic data operations, this paper achieves both.First identify the difficulties and potential security problems of directly with fully dynamic data updates from prior works. TPA can perform multiple auditing tasks simultaneously.

A uniform, integrated, machine-readable, semantic representation of cloud services, patterns, appliances and their compositions. Our approach aims at supporting the development of new applications for the Cloud environment,. Using semantic models and automatic reasoning to enhance potability and interoperability in multiple platform. Perform automatic discovery of Cloud services and Appliances; map between agnostic and vendor dependent Cloud Patterns and Services in semantic representations of cloud pattern. For future work, we plan to build a graphical tool for an easier, user-friendly categorization of Cloud services.

Easier querying, with automatic translation into our semantic machine-readable representation Used in multiple platform. Public-key encryption with keyword search (PEKS) is a versatile tool. It allows a third party knowing the search trapdoor of a keyword to search encrypted documents containing that keyword without decrypting the documents or knowing the keyword.A keyword privacy enhanced variant of PEKS referred to as public-key encryption with fuzzy keyword search (PEFKS).The search takes time linear in the size of the database storing by using of PEKS Schemes. It reduce the search time in both PEKS and PEFKS schemes.

RELATED WORK

Cloud deployment benefits based on the recent works addresses when to use the cloud services and when to migrate applications from the application provider. However these effects do not consider issue such as cost and consistency by using multipole cloud services like RACS, safe store,depsky. These system focus on issues pertaining to availability, durability, vendor lock. Unlike SPANStore none of the system are used to minimize the cost across the providers.

Optimizing cost and scalable storage of provisioning cost effective storage based on the work load characterization. To reduce a CPU overhead using minhash algorithm. In Storesim based on four main components that is deduplication based on SHA-1 algorithm, LMlayer based on BFS minhash and encryption decryption based on AES- 256. The storage depends on the bloom filter size in BFS minhash algorithm. It is very low and constant and increase the size of total data. . The additional aspects reinforce the logic aspect by enabling more pervasive inferences. As representative consider the first two services as in table Service able to discover the equivalences among these two services due to the differences of operation exposed and parameters exchanged. the context of a particular pattern but cannot be defined equivalent in general. These loose equivalent services can be exchanged without trouble of the services is used in the context of a particular pattern. In both existing PEKS schemes and our keyword privacy enhanced PEFKS scheme, the search takes time linear in the size of the database storing the keyword searchable ciphertexts. This is an affordable of the database if it is large. Hence, it is an interesting to reduce the search time in both PEKS and PEFKS schemes. We leave this for future work.

To ensure data security in cloud storage, it is critical to use a TPA technique to evaluate the service quality from an objective and independent perspective.The public auditability also allows clients to delegate the integrity of the verification tasks to TPA themselves can be unreliable or not be able to commit necessary computation resources performing continuous verifications. Another major concern is to construct verification protocols that can be used to accommodate dynamic data files. It explored the problem of providing simultaneousof the public auditability and data dynamics for remote data integrity check in Cloud Computing. Our construction is based on the deliberately designed to meet these two important goals with efficiency being kept closely in mind. To achieve efficient data dynamics, we have to improve the existing proof of storage models by manipulating the Hash Tree construction for block tag authentication. To support efficient handling of multiple auditing tasks, The further explore the technique of is based on the bilinear aggregate signature to extend our main result into the multiuser task setting, where TPA can perform multiple auditing tasks simultaneously. Extensive security and performance is used to analysis that the proposed scheme is highly efficient and provably secure.

CONCLUSION

A secure, efficient and dynamic search scheme is used to propose,that does not supports only the accurate multi keyword ranked search but also the dynamic deletion and insertion of documents can be performed using it. We construct a special keyword balanced binary tree as the index, and to obtain better efficiency than linear search. We proposed a hybrid scheme that combines public key encryption and fully homomorphic encryption. The proposed scheme is suitable for cloud computing environments since it has low storage requirement, and supports efficient computing on encrypted data.Our solution provides the size of the transmitted ciphertexts and the conversion. The parameters of our hybrid scheme are very large when the message space of the FHE. StoreSim aims to store syntactically similar data on the same cloud, thus minimizing the user's information leakage across multiple clouds.

REFERENCES

[1] Ning Cao, Cong Wang, Ming Li, Kui Ren, Wenjing Lou
"Privacy-Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data" IEEE 2017.
[2] Qingji Zheng, Shouhuai Xu, Giuseppe Ateniese "VABKS: Verifiable Attribute based Keyword Search over Outsourced Encrypted Data IEEE 2016.
[3] Zhe Wu,Michael Butkiewicz,Dorian Perkins" SPANStore:Cost-Effective Geo-Replicated Storage Spanning Multiple Cloud Services" IEEE 2015.
[4] T.G. Papaionnou,N.Borvin,K.Aberer IEEE "SCALIA: An Adaptive Scheme For Effective Multi-Cloud Storage"2012.
[5] Yanjiang Yang, Haibing Lu, Jian Weng "Multi-User Private Keyword Search for Cloud Computing" IEEE 2013.
[6] Zhihua Xia, Member, Xinhui Wang, Xingming Sun,Wang,Qian Wang "A Secure and Dynamic Multi-Keyword Ranked Search Scheme over Encrypted Cloud Data" IEEE Transactions on Parallel And Distributed Systems Vol. 27 February 2012.60
[7] Wenhai Sun, Shucheng Yu, Wenjing Lou, Y. Thomas Hou, and Hui Li Xidian University, Xi'an, Shaanxi "Protecting Your Right: Attribute -based Keyword Search with Fine-grained Owner-enforced Search Authorization in the Cloud"IEEE 2011.
[8] Qian Wang, Cong Wang, Kui Ren, Wenjing Lou, and Jin Li"Enabling Public Auditability and Data Dynamics for Storage Security in Cloud Computing"IEEE Transactions vol. 22, May 2011.
[9] Peng Xu, Hai Jin, Qianhong Wu, and Wei Wang "Public-Key Encryption with Fuzzy KeywordSearch: A Provably Secure Scheme Under Keyword Guessing Attack"IEEE 2011.
[10] Beniamino Di Martino, Antonio Esposito and Giuseppina Cretella "Semantic Representation of Cloud Patterns and Services with Automated Reasoning to support Cloud Application Portability"IEEE 2010.