International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue VIII, August 2018- Available at www.ijraset.com

Vital Investigation of Data in Hybrid Cloud

K Pavan[1], N Nalini Krupa[2], N Madhuri[3]
[1, 2, 3]*Computer Science And Engineering*

Abstract: ***In terms of computing, data Is essential however by the inclusion of redundant data may consume much storage space. To save the bandwidth we must remove the data which is the replicas of original, to achieve this we have so many techniques of compressing data in that deduplication is one. It is important to have the protection to maintain the secrecy of the sensitive data of the peer groups. To proceed with this, we must maintain some cryptographic standards. In our paper we are implementing a typical approach that must provide security to our data and must eliminate the redundant copies of the data. To achieve this, we are going to scan the copies of the data and another main issue associated with this is the security to go through with some convergent cryptographic techniques. All in all, this approach is going to give the proof of idea on various security investigations of our definitions. Also, we execute a model of copy check &leads some successful results.***
Key words: token, s-csp, public cloud, sensitive, confidential

I. INTRODUCTION

Cloud computing is one of the emerging technologies in today's modern computing. It provides an illusion of accessing infinite amount of resources to the user by hiding the accessing mechanisms as a service over internet. These can provide high quality storage space, accurate and reliable computing results at an affordable price. As it is having much fame user also willing to place larger data on it also can manage and share the data to the other users. While providing the unlimited high-quality storage to the user the first question rises weather is it able to manage that, large volumes of data? To achieve this, we need to have efficient management mechanisms.

Without wasting the storage space of abandon space, we have well known and simple de-duplication techniques to achieve it, which is most popular in recent times and gets the most concentration of the users of storage management.

It is the popular compression technique to remove the replicas of the content stored in the cloud. The basic mechanism why we elect this is, while participating in the communication with the network there is a chance to do interchange of data from the user to data repository, the integration of multiple number of bytes transferred to the repository will increase the utilization of the data store, our motto is also the same i.e. to improve the utilization of the data store management.

Unfortunately, somehow, we are storing the replicas of original contents in the storage space, by implementing the de-duplication mechanism which eliminates the duplicate copies of the data. It is also capable of checking the de-duplication in block level as well file level. If it is the case of block level it can check the block of data that is generated in the dissimilar bunch of files. If it is the case of file level it eliminates the replicas of the same file name or contents.

De-duplication of data provides so many advantages, but it can also get some flaws, that the data of the users is able to get attack in terms of security or private concerns. Past security mechanisms provide secrecy to the data but there is no de-duplication and compatibility. In traditional standards of cryptography data encryption and decryption is done through keys of you want to encrypt the data you need another user to do it. Then there is a possibility to place multiple cipher texts in the cloud, which making the space with redundant copies of information to the plain original text by having in multiple places.

The other is convergent mechanism which can achieve data confidentiality. It can encrypt and decrypt the data based on some cryptographic hash values given to perform computing on the data. By performing key generation to encrypt the data user preserves it and place the cipher text in the cloud. Encryption is a finite operation, it is done from the data, same key is generated to the same data which results same cipher texts. To restrict the access, a protocol is needed as a proof which was detected by the own or same file. After placing this no other user is able to place the same file into the store, the server won't give the permission to do. Later the user who wants to decrypt the data, the key which is given by the owner of the data is used and can be downloaded from the server then performs the decryption of the cipher text.

II. LITERATURE SURVEY

1) Distributed file system allows replicating the files over multiple computers, this uses significantly vast storage space and gets back the used space is crucial thing. J.R.Douceur [2] proposes a method to get the space over duplication for controlled file replication. It even includes convergent security measures with multiple keys of different users.

2) Bellare .et. al. proposes a mechanism to copy a file into the cloud and run for duplicate check to save the storage space, to do that we have to encrypt the file there we lost the space of our savings by having multiple cipher texts. Even more it is favorable to brute force attacks used to recover the lost files. To avoid this, they proposed a mechanism called DupLESS which encrypts the contents based on PRF protocol which the client is not aware of it. There by altering the predictable to unpredictable content of the message.

3) Li.et. al [6] proposes a technique to integrate the space of storage by eliminating duplicates and to improve the upload bandwidth. Even convergent encryption yields safe duplications, it has to manage a lot of user keys which is a crucial point to organize. Our analyses in terms of security gives better over, only definitions as a proof to this we must use LSD method which shows the overheads in real world.

4) Stanek. et. Al[5] past five years have shown us the fashion of services related to cloud for big volumes of data processing and managing. Over all security & privacy are on the apex with cloud structures. By using open stack swift a new storing and sharing mechanism is implemented as a two- fold process. A stronger security mechanism is used to the sensitive data of the user, which should get efficiency over the data. Sometimes they may fail to maintain data with the keys associated with them.

III. SYSTEM ARCHITECTURE

Our Architecture is to be treated as a hybrid which consists of both public & private clouds, in which each of them are included to get the result in a proper and accurate way. Mainly the public cloud is becoming the cloud where is our actual data is stored in the encrypted form with which we are accessed by some permissions and keys of the owner, which we can't do any modification on it. Also, it is acted as the data organizer.

On the other hand, the private cloud which is acted as the key organizer where we kept all the users with the associated token ID's of the individual file requests or responses. The mechanism is very simple while we want to perform any operation on the data in the public cloud that is, mapped with the privileged key, on the private cloud. Without having the key info user can't able to access the data of their own.

If a user wants to perform an operation to be as retrieving the content or storing the content, he/she must acquire key which is treating as the token. Only by having this he/she can do the operation intern; public cloud is synchronized with the keys of private cloud. Once he /she get the token then only he can do the operation which is required. Whatever be the operation he must hold his/her keys.

All the keys are mapped to the service provider. By the involvement of him only the modifications to the individual users are going to be happening. He can also able to maintain the session, cancel the permissions of different users etc.

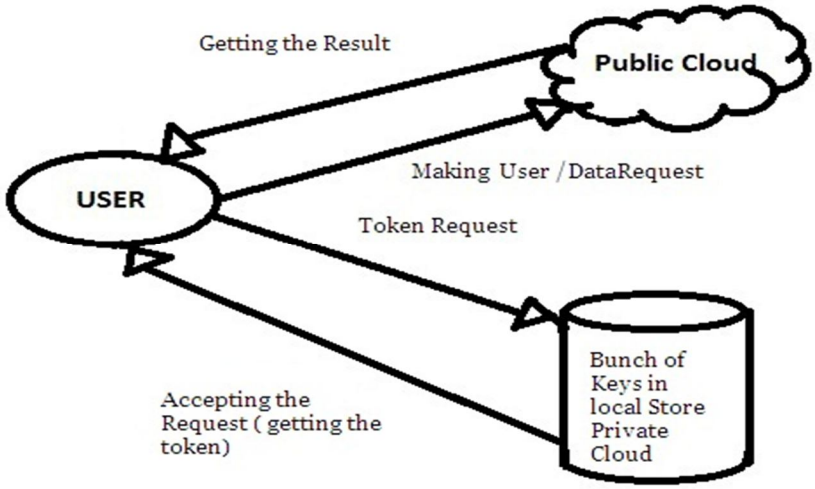

Fig: Basic structure of the model

IV. PROPOSED APPROACH

In this we are implementing three modules each are having their specialty. Those are

A. Secure Cloud Service Provider Module

This is one of the essential modules, which deals with the service of storing data in public cloud. It provides the facility to outsource the data as well the services to the users.

To slash the cost of storage this module is going to eliminate the storage of redundant data and keeps only the genuine copy by the technique of duplication. To achieve this, we insist this should always be in online and having high computational capacity with unrestricted storage capability.

B. Data User Module

Any of the streams the users are in the extreme end level to utilize the data or services in the hierarchy of usage. Even in this the user wants to process the data to store in the cloud and then access it. To save the bandwidth we are storing only the genuine copies of data, which may be of single or group of users. In the reliable de-duplication system every user is associated with some tokens, to perform some operations with the files in the cloud. Each file is getting protected with their associated keys within the allocated privileges by the super user and is stored in the private cloud.

C. Private Cloud Module

This is the key module of our implementation, where all the privileges of the files of individual users are placed. In our mechanism without having a key user won't perform any operation with the file. To provide the flexibility to the user that his/her file stored in cloud. We must use this phase apart from traditional techniques.

In general, users are restricted version of using files of computing resources on the other hand our public cloud is not fully secured, to overcome these two flaws we have this private cloud module where it is creating interface between the users and contents of cloud storage.

All the keys are managed by this private cloud to organize and give the response to the requests of the users. The service of this private cloud module is to make the secure requests to the cloud &get the results in uncorrupted way.

1) *Secure De-Duplication System:* We are having several types of protection techniques relevant with privacy, in that correctness of duplicate token check is widely used one. In this we have external and internal adversaries without getting any privilege external adversary can be internal. Without forging, a user requires a privilege and able to get a valid duplicate token with any other privilege on the file here both privileges won't match. It is important to have a token without making the request from its own private cloud server; it can't forge and give the valid duplicate token that has been asked for file.

V. EVALUATION OF PROPOSED APPROACH OVER TRADITIONAL

Here we are going to compare the performance how it should get the data by eliminating the replicas of the data in the cloud storage. It is based on the ratio of the de-duplication. To achieve this, we are implementing the check on a text file of the size 200KB in multiples varies from 20MB to 50MB. We can prepare a couple of sets one includes the original and unique copy of data and other with duplicate copies of the data.

Here the upload and encryption are done in the files of unique one and the other set is skipped by having the duplicate files, what we said in our proposed approach. It also decreases the amount of time to check for duplication. The replicas are found in our file it will integrate the time required to complete the duplicate check, on an average for a file with duplications the rate will be only 30% with the corresponding file.

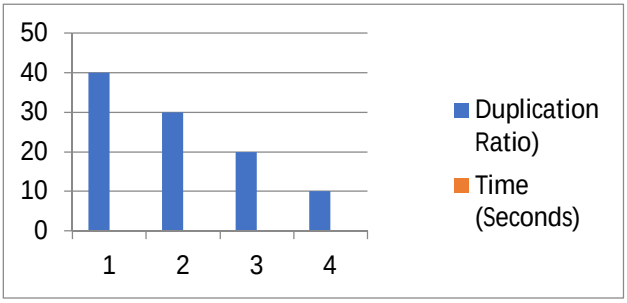

Fig: Comparison chart

VI. CONCLUSION

Our focus in this paper is to analyze the problems associated with the duplication in the computation process of the data in the cloud. Not as in traditional approaches inclusion of private cloud makes much sense and act as the interface to the owners of the data and to the public loud in a convenient way. This kind of system structure grabs the attention of researchers. While performing the operation on the data only private cloud to be active without that owner won't able to manage the data in the public cloud, because of this our system is supporting hybrid cloud architecture which consists of both public cloud for the data and private cloud for the keys. Users can perform the upload only after the checking of duplication with the respected mechanism, once it is done everything should be like normal.

REFERENCES

[1] J.R. Douceur File level and block level detection
[2] Bellare .et. al. Secure auditing and deduplication