International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 02 Issue: 06 | Sep-2015 www.irjet.net p-ISSN: 2395-0072

A smart data backup technique for cloud computing using seed block algorithm strategy

Mr. G. S. Narke[1]
Comp. Dept. BVCOERI
Nasik, India

Mr. A. K. Harijan[2]
Comp. Dept. BVCOERI
Nasik, India

Mr. A. R. Shinde[3]
Comp. Dept. BVCOERI
Nasik, India

Prof.H.D.Sonawane[4]
Comp. Dept. BVCOERI
Nasik, India

Abstract - Large amount of data is generated in cloud computing, data in electronic form. The data recovery services are required to maintain this data very efficiently, In this paper we have proposed a smart remote data backup algorithm, Seed Block Algorithm (SBA). The main aim of proposed algorithm is two types, that will first one is it will help the users to collect information from any remote location in the non presence of or loss of network connectivity. Another one is to recover the files if by mistake file gets deleted or if the cloud gets destroyed due to any reason. The time related problems are also solved by proposed SBA such that it will take minimum time for the recovery procedure. Without using any of the existing encryption techniques, the proposed paper concentrates on security features for the back up files stored at remote servers.

Key Words: *Central Repository, Remote Repository, Parity Cloud Service, Seed Block, etc..*

1. Introdution

Indian National Institute of Standard and Technology of India defines as a model for enabling convenient, on-demand network access to a share combination of configurable computing services (for ex- networks, servers, storage, applications and services) that can be provided fastly and released with minimum management effort or services provider . In today's world, Cloud Computing is itself a very huge technology which is passing all the previous technology of computing (like cluster, distributed etc.) of this combative and challenging IT world.

As the advantages of cloud computing overcoming that disadvantage of various early computing techniques, so the need of cloud computing has already started.

Cloud storage provides online storage facility where the is data stored in form of virtualized pool that is mostly hosted by third parties.

The hosting companies operates large data on large data center and according to the need of the customer these data center visualized the resources and disclose them as the storage pools that will help user to store files or data objects. As lots of user shares the storage and other resources, it may happen that other customers can access your data. Either the human mistakes, faulty devices, network connectivity, a bug or any criminal intention may harm or put our cloud storage on the risk or danger. And changes in the cloud are also made very frequently; we can name it as dynamic data. Various operations are supported to this dynamic data as like insertion, deletion and block modification, updation.

Since the services are not only limited for archiving and taking backup of data; remote data integrity is also needed. Because the reason is that the data integrity always aims to focus on the validity and fidelity of the entire state of the server that takes good care of the heavily generated data which remains same at the time of storing at primary cloud remote server and transmission. Integrity plays vital role in back-up and recovery services. In Cloud computing most of the new recovery and backup methods have developed. This will we study in our literature like , Cold/Hot backup strategy [1], PCS[2], Linux Box [3], etc. that, discussed the data recovery process. However, there is still various successful techniques which are lagging behind some critical issues like low cost, implementation complexity, process security and time related issues.

To cater this issues, in this paper we propose Seed Block Algorithm (SBA) which is one of the smart data backup algorithm for remote data access. The contribution of the proposed SBA is twofold;

1-SBA helps the users to collect information from any remote area in the failure of network connection.

2-Recover the files in case if it gets deleted due to any reason like by mistake or intentionally or if the cloud gets destroyed.

This paper is organized as follows: Section II focuses on the related literature of existing methods that are successful to some level in the cloud computing domain. In Section III, we discuss about the remote data backup server. Section IV describes the brief description of the proposed seed block algorithm (SBA) and Section V shows the results and demonstration analysis of the proposed SBA. Finally, in Section VI conclusions are given.

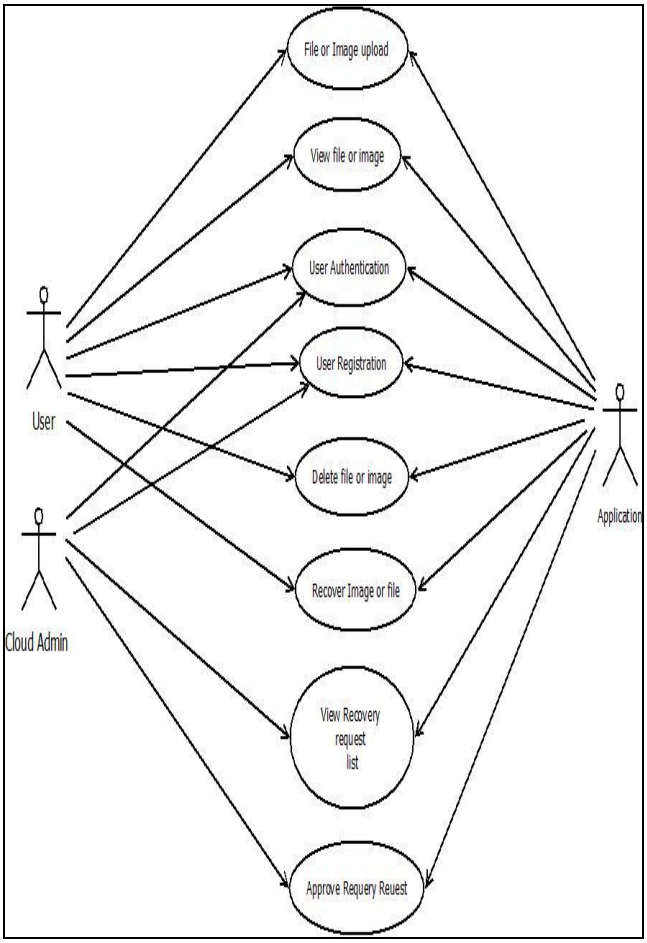


Fig. Activity Diagram

2. Releted literature

There are various cloud computing domain like cold and hot backup techniques, linux box techniques, etc. These recent recovery techniques that had been developed , will study in our literature. Under uncontrolled circumstances like low implementation complexity, file recovery, system cost, redundancy performance of these techniques is degraded. One of these technique is known as PCS technique is among above techniques is reliable. It is easy and simple to use for recovery of data. This technique is based on parity recovery service. It have high probability of data recovery. Exclusive-OR is used by PCS for recovery of data using virtual disk but it can not control complexity implementation .

For mobile users, such as smart phone and laptop users another very efficient technique is developed which is known as HS-DRT technique. For recovery HS-DRT technique requires high cost and not able to data duplication control. Data transfer using distributed mechanism and encryption technology with high speed, that type of technology is known as innovative backup method . HS-DRT does data transfer using distributed mechanism and encryption technology with high speed. Therefore, HS-DRT technique is called as innovative backup method . Two sequences are proposed by HS-DRT technique these two sequences are as follows :

1-Backup sequence

2-Recovery sequence

In that, gathering of data and backup it is the work of first one known as backup sequence. if any calamities is happened by natural like earthquake , flood, fire etc. or man-made, recovery of the information is the task of recovery sequence . There are various components of HS-DRT among which is starting of recovery sequence is the objective of supervisory server, components of HS-DRT. There also limitation to the HS-DRT model so that this is not a perfect solution to our problem.

Another model is known as Linux box model, it is very simple recovery & data backup concept using minimum cost. As there is minimum cost therefore security level is low. It also transfer the control from one cloud to another cloud very easily, this is one of the advantage of linux box. Other advantages are like the customers who have small or medium business can afford this model. The ISP dependency and required backup cost can be minimized using this solution. Simple linux box works at file level which is provided by cloud service provider to user. Cloud to local drive backup is performed by an application in linux box. The linux box provides encrypted also secure transmission . Entire virtual machine synchronization can be done using linux box model, which indicates bandwidth waste. As like that router failure scenario and significantly cost reduction can be achieved using SBBR . If router failure occurs there is no change in IP connectivity, SBBR concerns with it. In multilayer signaling fundamental services like network management is provided by SBBR.

Literature shows that how max outer requirements are imposed by specified services. For setting Seed Block Requirements Architecture (SBRA),which is directly affecting by outsage requirement of specified services. There is not possible to introduce optimization techniques required for minimization of code along with cost reduction.

Therefore, we have use entirely recent concept/technique known as Virtualization REN Cloud, which focuses on low cost implementation for its infrastructure. There is another model, which also using for cost reduction service commonly recognized as Rent Out Rented Resource(RORR). Rent Out Rented Resource made up of three phases commonly known as

1.Discovery

2.Matchmaking

3.Authentication

We use all above techniques for reduction of implementation cost. In Cold and Hot techniques, as data increases gradually Service's cost also increases. Service's failure is detected by Cold Backup Replacement strategy used for process recovery and Dynamic network service composition is archived using Hot Backup replacement

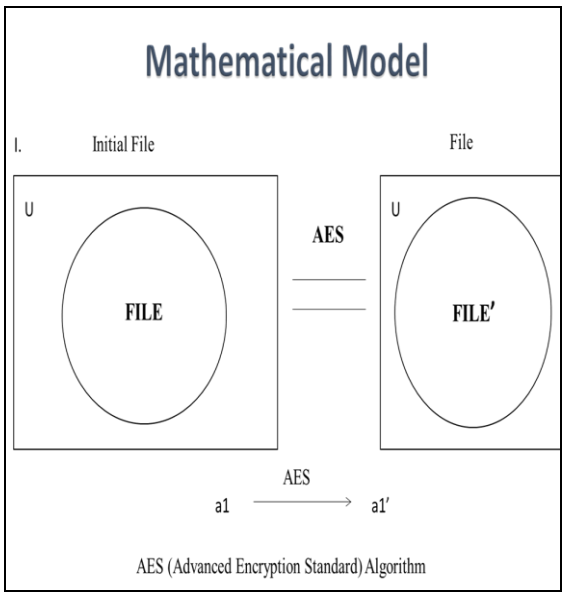
strategy for transcendental recovery. There is always activated states for backup services. Service composition's successful implementation is the first result which will we get.

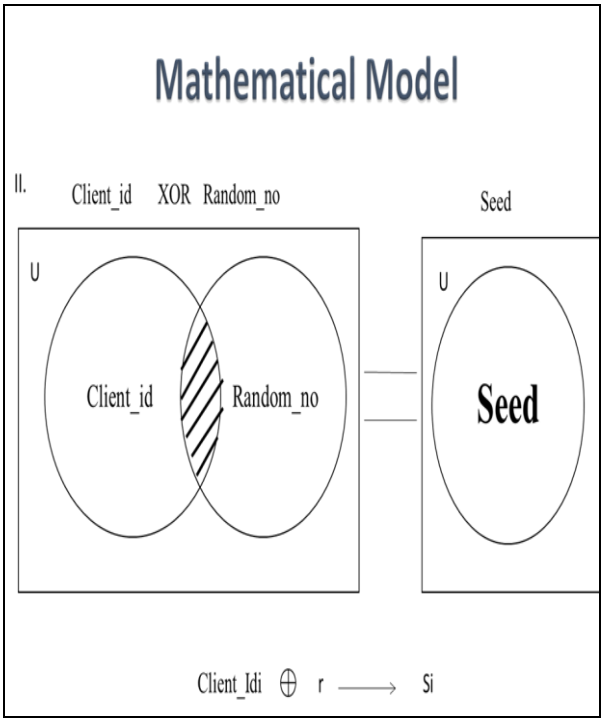
Although, in every solution of backup techniques, not able to gain all problems of remote backup of data server. Backup processes are highly accepted in industrial area because these backup techniques are emerging research topic and also very crucial.

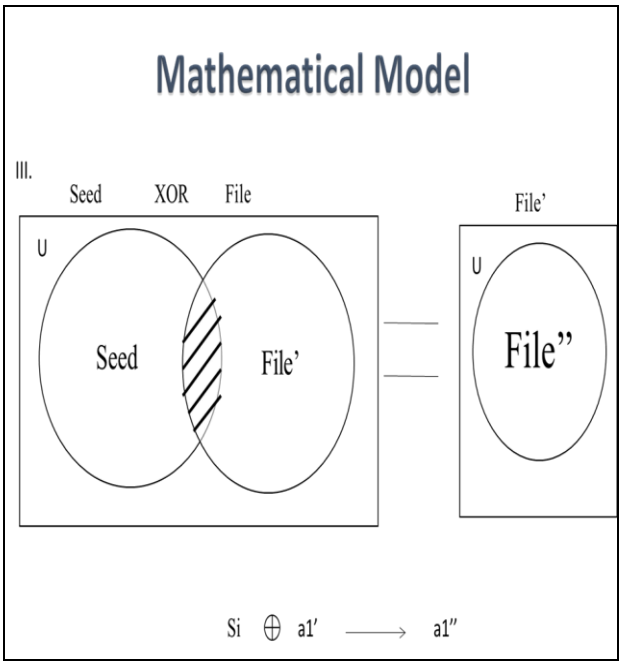
3. Problem Statement

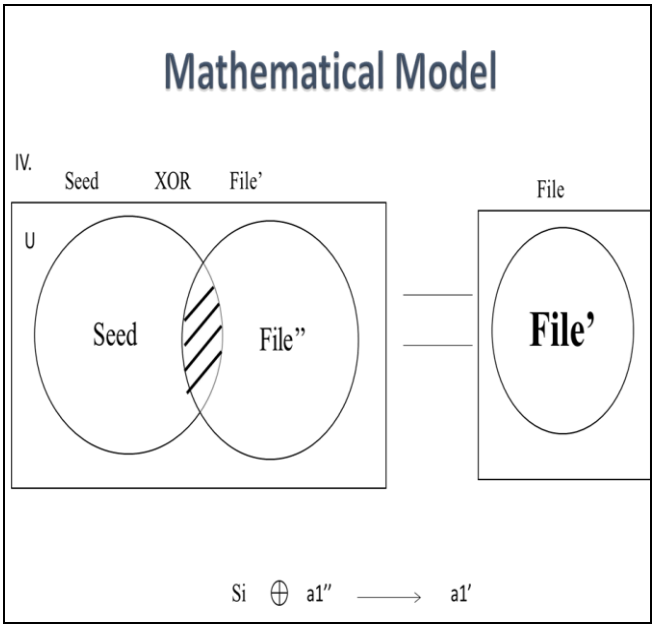
Data recovery services must be required to maintaining Data efficiency at the server. We are proposing a very efficient algorithm for data backup in this paper, this algorithm is called as Seed Block Algorithm (SBA). In Seed Block Algorithm, there are two types of objectives which we are trying to achieve , these are as : First one is from any Distinct location user collecting information when network connectivity is absent. This algorithm is helping to done this. Another one is if file deletion is occurred due to any reason or bad luckily cloud gets destroyed then still we can recover that certain data.

4. Mathematical Module Diagram:









5. Existing System

Vijaykumar Javaraiah [4] Presented that, people thinks that, copy of Primary cloud is Primary cloud backup server. When that backup server presents far away from the user that is remote location from the Primary server. Primary cloud's complete states are contained by backup server , then this remote location server is named as Remote Data Backup Server. The central repository is also recognized as Main cloud of the system. The Remote repository is recognized as Remote Backup of the system.
Sheheryar Malik, Fabrice Huet, December 2011[9] in thaair work, there are natural calamity like earthquake, flood, fire etc. in such circumstances Central also known as main repository can lost its information or it can be due to other circumstances like accident or by mistake deletion of that information or by any Human attack and then it uses the information from the remote repository. From remote repository user collecting the data or information. To help user for doing that task is the Remote Backup facility's main aim even if server is not able to connect the network or if data not available on primary cloud. As shown in Fig-1 If the information is not available on remote repository still client can have the permission for accessing that file on Remote repository(i.e. indirectly). The following issues must be covered in Remote Backup services:

1) Data Integrity
Server's whole structure along with all complete states tells us about Data integrity of server. At the time of transmission and reception, data which resist to any kind of change in it. Such type of data is verifies using Data Integrity. Validity of Data on Remote server is also checked by Data integrity.

2) Data security
The Remote server have primary priority to provide total security to data of user. And either intentionally or non intentionally, only particular user should have access to that data or not any other users.

3) Data Confidentiality
In certain times, we have to kept user's data files to be secret such that if no. of users simultaneously accessing the cloud, when other users accessing files on the cloud should unable to see particular data file that is belongs to only that particular user. This is also known as Data Confidentiality characteristic.

4) Genuine Characteristic
Trustworthiness is the important characteristic of the Remote cloud. Remote cloud should posses that because every user having their private also confidential data on cloud. Therefore Trustworthiness characteristic should be present in remote also in cloud backup.

5) Cost efficiency
The cost of processing of data recovery should be efficient so that large number of companies along with users can take benefit of back-up and recovery service. There are many large numbers of methods that have focused on these issues. The foresaid issues occurs at the time of recovery also in back-up of domain of cloud computing are discuss in following techniques later.

6. Proposed system

In literature we have studied many techniques for backup and mechanism purpose like HSDRT[1], PCS[2], Linux Box[3],Cold and Hot backup[4] etc. In cloud computing problems like less economical, risk and time related problems are still not solved which we have discussed previously.
As discussed above low implementation complexity, low cost, security and timing related issues are still challenging in the field of cloud computing. To solve these issues we propose SBA algorithm and in next section, we will discuss the design of proposed SBA in detail.

Seed Block Algorithm (SBA) Architecture

This algorithm focus on simple mechanism of the back-up and recovery process. It mainly uses the idea of Exclusive–OR (XOR) method of the computing world.

E.g-: - If suppose there are two data files: P and Q. When we XOR P and Q it produced X i.e.
X =P Ex-OR Q .
If suppose P data file get deleted and we want our P data file back then we are able to get P data file back, then it is very easy to get back it with the help of Q and X data file .i.e. P = X Ex-OR Q.
Alike, the Seed Block Algorithm works to provide the simple Back-up and recovery process. Its architecture is

shown in Fig-2 which includes the Main Cloud and its clients and the Remote Server. Here In the cloud computing we have to set any kind of number and client id which must be unique. Second, In the main cloud whenever the client id is in registration phase; at that time client id and a kind of that random number is getting EXORed ( ) with each other to produce seed block for the relevant client. The generated seed block relates to each client is stored at remote server.

Whenever client will create the file in cloud first time, it is stored at the main cloud. When it is stored in main server, the main file of client is currently EXORing with the Seed Block of the particular client. In remote server that resulting EXORed file is stored in file format. If by mistake file in main cloud gets crashed / damaged or file is gets deleted unexpectedly, then in this situation user will retrieve the original file by EXORing file' with the seed block of the relevant client to generate the original file and return the resulting file i.e. original file back to the requested client.

In this section, we discuss the demonstration and result analysis of the SBA algorithm. For demonstration we focused on different minimum system requirements for main cloud's server and remote server. Memory requirement is kept 8GB for the main cloud's server and 12GB and remote server which can be extended as per the requirement. it is observed that memory requirement is greater than the main cloud's server due to additional information is kept onto remote server E.g- different Seed Blocks of the corresponding client .

During demonstration we found that the size of native data file stored at main cloud is same as the size of Back-up file stored at Remote Server. We perform this demonstration for different types of files. SBA is very much Vigorous in maintaining the size of recovery file same as that the original data file as results for this experiment showed us. From this we can conclude that proposed SBA recover the data file without any loss of data.

Time taken by the process [processing time] when client uploads a file at main cloud and that includes the assembling of data like any random number from main cloud, seed block of the corresponding client from the remote server for EX-ORing purpose; after assembling, performing the EXORed operation of the contents of the uploaded file with the seed block and at the end store the EXORed file on the remote server. We also examined that as the data size increases, the processing time also increases. By the way, we also noticed that performance which is megabyte per sec (MB/sec) is constant at some extent even if the data size increases.

Main Cloud's CPU usage starts with 0% and as the client uploads the file onto it then usage increases; such that it has to check whether the client is authenticated or not, at the same time it sends request to Remote Server for the relevant Seed Block. When request reaches to Remote Server it starts collecting the details as well as the seed Block and gives reply in form of the seed Block and during this session, load at Main Cloud decreases which in turn causes for low CPU usage at main cloud. After receiving the requested data, at primary cloud CPU usage level increases due to EX-ORed operation process. Again the Final EX-ORed file sends to Remote Server.

ADVANTAGES OF PROPOSED SYSTEM

1. Recovery of same sized data
2. Privacy is the primary objective
3. low cost implementation

7. Flow diagram

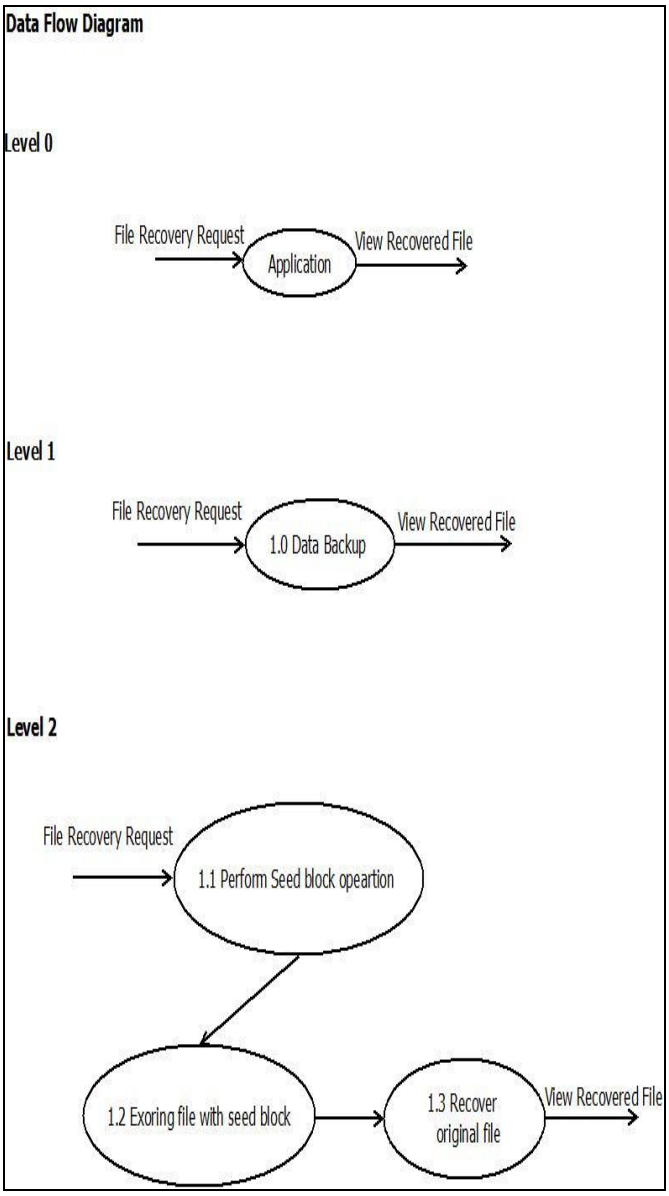


FIG. FLOW DIAGRAM

8. Algorithm used

*1. **Start***
2. first we set a random number in the main storage and unique client id for every client
3. Whenever the client id is being register in the main storage,
then client id and random number is getting EXORed ( ) with each other to generate seed block for the particular client.
4. Whenever client creates the file in cloud first time, it is stored at the main storage.
5. When it is stored in main storage (blob), the main file of client is being EXORed with the Seed Block of the particular client.
6. It is also encrypted using public key RSA
7. That output file is stored at the backup storage (blob) in the form of file' (pronounced as File dash).
8. During Retrieval, check if data present in main storage
If present then EXOR with seed block and retrieve data
If not present, retrieve data from backup storage
9. During Retrieval from backup storage , the private key of the user will used to decrypt file'
10. The user will get the original file by EXORing on decrypted file' with the seed block of the corresponding client to produce the original file and return the resulted file in case of crash.
11. Stop

9. Module Information Diagram:

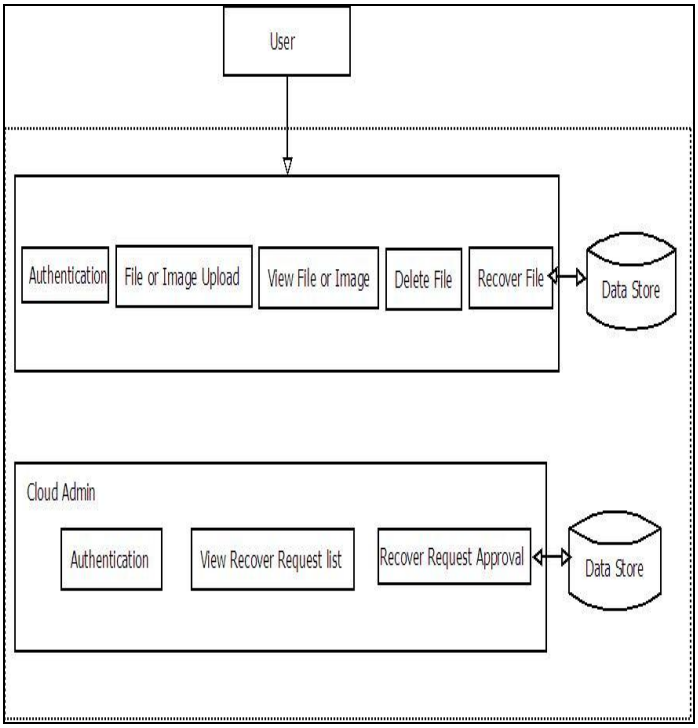


***FIG**.SYSTEM ARCHITETURE*

10. Conclusion:

In this paper, we have presented detailed design of proposed SBA algorithm. Proposed SBA is robust in assisting the users to collect information from any remote location in the loss of network connection and if file deletion occurs due to any reason or bad luckily cloud gets destroyed till we can also recover files. Demonstration and result analysis shows that proposed SBA also focuses on the security issue for the back-up files stored at remote server, without using any of the existing encryption techniques. The proposed SBA will take minimum time for process recovery so that the issues corresponds to time can be solved.

11. Acknowledgment:

For the valuable guidance, continues support and inspiration of my respected sir Mr. H. D. Sonawane, Computer Engineering, BVCOE&RI, Nashik, It is my pleasure to express my knowledge to him. This paper could not be success without apps analysis done which help to understand for this paper necessity and also thanks to Ms.Kruti Sharma and Prof.K.R.Singh for their valuable help and guidance.

12. REFERENCES

[1] Chi-won Song, Sungmin Park, Dong-wook Kim, Sooyong Kang, 2011, "Parity Cloud Service: A Privacy-Protected Personal Data Recovery Service," International Joint Conference of IEEE TrustCom-11/IEEE ICESS-11/FCST-11.

[2] Y.Ueno, N.Miyaho, and S.Suzuki, , 2009, "Disaster Recovery Mechanism using Widely Distributed Networking and Secure Metadata Handling Technology", Proceedings of the 4th edition of the UPGRADE-CN workshop, pp. 45-48.

[3] Giuseppe Pirr´o, Paolo Trunfio , Domenico Talia, Paolo Missier and Carole Goble, 2010, "ERGOT: A Semantic-based System for Service Discovery in Distributed Infrastructures," 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing.

[4] Vijaykumar Javaraiah Brocade Advanced Networks and Telecommunication Systems (ANTS), 2011, "Backup for Cloud and Disaster Recovery for Consumers and SMBs," IEEE 5th International Conference, 2011.

[5] S. Zhang, X. Chen, and X. Huo, 2010, "Cloud Computing Research and Development Trend," IEEE Second International Conference on Future Networks, pp. 93-97.

[6] T. M. Coughlin and S. L. Linfoot, 2010, "A Novel Taxonomy for Consumer Metadata," IEEE ICCE Conference.

[7] K. Keahey, M. Tsugawa, A. Matsunaga, J. Fortes, 2009, "Sky Computing", IEEE Journal of Internet Computing, vol. 13, pp. 43-51.

[8] M. D. Assuncao, A.Costanzo and R. Buyya, 2009, "Evaluating the Cost- Benefit of Using Cloud Computing to Extend the Capacity of Clusters," Proceedings of the 18th International Symposium on High Performance Distributed Computing (HPDC 2009), Germany.

[9] Sheheryar Malik, Fabrice Huet, December 2011, "Virtual Cloud: Rent Out the Rented Resources," 6th International Conference on Internet Technology and Secure Transactions,11-14 ,Abu Dhabi, United Arab Emirates.

[10] Wayne A. Jansen, 2011, "Cloud Hooks: Security and Privacy Issues in Cloud Computing, 44th Hawaii International Conference on System Sciences.Hawaii.