Data Recovery Technique using Seed Block Algorithm for Cloud Computing

J.Sangeetha Priya[1] M.I Asifa Saman[2] S.Neevedha[3] V.Suganya[4] M.J Zainiya Nazrin[5]
[1,2,3,4,5]Department of Information Technology
[1,2,3,4,5]Saranathan College of Engineering, Trichy-620 012, Tamil Nadu, India

Abstract— A large amount of electronic data is generated in cloud computing. In order to maintain the data efficiently data recovery services are used. To tackle this in this project we use a Data Backup and Recovery using seed Block algorithm. The objective of this algorithm is in two phases. First it helps the user to collect information from any remote location in the absence of network connectivity and then secondly it recovers the files in case if the file is deleted or if the cloud gets destroyed due to any reason. Seed Block Algorithm takes minimum time for recovery. It also provides security for the backup files that are stored at the remote server without the use of any existing encryption technique.

Key words: Proxy, cloud service, Seed Block, Main Cloud, Backup Cloud

I. INTRODUCTION

A. Basics & Essentials Related to Cloud Computing:

1) Cloud Computing:

Cloud computing is the delivery of computing as a service rather than a product, whereby shared resources, software, and information are provided to computers and other devices as a utility over a network. Cloud computing is Internet-based computing, whereby shared servers provide resources, software, and data to computers and other devices on demand, as with the electricity grid.

The term "cloud", to have its origins in network diagrams that represented the internet, or various parts of it, as schematic clouds. Cloud computing refers to the many different types of services and applications being delivered in the internet cloud, and the fact that, in many cases, the devices used to access these services and applications do not require any special applications.

Cloud computing is a technology that uses the internet and central remote servers to maintain data and applications. Cloud computing allows consumers and businesses to use applications without installation and access their personal files at any computer with internet access. This technology allows for much more inefficient computing by centralizing storage, memory, processing and bandwidth.

A simple example of cloud computing is Yahoo email, Gmail, or Hotmail etc. You don't need software or a server to use them. All a consumer would need is just an internet connection and you can start sending emails. The server and email management software is all on the cloud (internet) and is totally managed by the cloud service provider Yahoo, Google etc. The consumer gets to use the software alone and enjoy the benefits.

2) Cloud Computing Terminology:

- While the debate on the actual definition of cloud computing rages on, it seems that a whole new cloud computing vocabulary is rapidly emerging. I thought I'd list some of the new terms I'm seeing with brief definitions, examples of usage and references to discussions related to these terms. Hope this is useful.
- Cloudburst: The term cloudburst is being use in two meanings, negative and positive: Cloud computing infrastructure is one of the terminologies of cloud hosting; it is actually software which can run through the internet. This infrastructure is known to share resources and information to different devices, like computers and electricity grid.
- The industry of IT is being supplied by Linux VPS hosting and cloud computing with models that are quite delivery and supplementing. Some people are worried because of the cloud computing costs, but everything will surely be worth it. Cloud hosting is not actually for everyone who owns a network. This is usually for businesses that need flexibility without having so much provision, so if you only have a small site then I guess you no longer need this cloud hosting. Actually businesses who want to grow are the ones who are availing of this hosting especially that the cloud computing pricing is a bit high.
- For people who need a desktop that is virtual, then the cloud computing OS is that they need. It is a Cloud Computing Platform web that will give people a desktop that is virtual. This is ideal for people who have businesses on the internet, and who are always travelling. You don't need to bring your computer anywhere you go, because your virtual desktop can be access no matter where you are in this world. If you think that you will still need a remote computer for you to be able to access your desktop, then you are wrong because you can simply access it without your remote.
- If you want your business to grow perfectly and positively, then I suggest that you get your own cloud hosting and computing, because I assure you that everything you need will be given by these two. Stop having second thoughts, about buying these things because it will surely be worth it; imagine having something that will help your business.

3) Secure Architecture Models:

Open Security Architecture (OSA) provides free frameworks that are easily integrated in applications, for the security architecture community. Its patterns are based on schematics that show the information traffic flow for a particular implementation as well as policies implemented at each step for security reasons.

4) End Users:

End Users need to access certain resources in the cloud and should be aware of access agreements such as acceptable use or conflict of interest. In this model, end user signatures may be used to confirm someone is committed to such policies. The client organization should run mechanisms to detect vulnerable code or protocols at entry points such as

firewalls, servers, or mobile devices and upload patches on the local systems as soon as they are found. Thus, this approach ensures security on the end users and on the cloud alike. However, the cloud needs to be secure from any user with malicious intent that may attempt to gain access to information or shut down a service. For this reason, the cloud should include a denial of service (DOS) protection. One way of enforcing DOS protection is done by improving the infrastructure with more bandwidth and better computational power which the cloud has abundantly. However, in the more traditional sense, it involves filtering certain packets that have similar IP source addresses or server requests. The next issue concerning the cloud provider to end users is transmission integrity. One way of implementing integrity is by using secure socket layer (SSL) or transport layer security (TLS) to ensure that the sessions are not being altered by a man in the middle attack. At a lower level, the network can be made secure by the use of secure internet protocol (IPsec). Lastly, the final middle point between end users and the cloud is transmission confidentiality or the guarantee that no one is listening on the conversation between authenticated users and the cloud. The same mechanisms mentioned above can also guarantee confidentiality.

5) System Architects:

System architects are employed with writing the policies that pertain to the installation and configuration of hardware components such as firewalls, servers, routers, and software such as operating systems, thin clients, etc. They designate control protocols to direct the information flow within the cloud such as router update/queuing protocols, proxy server configurations or encrypted tunnels.

6) Layers in Cloud Computing:

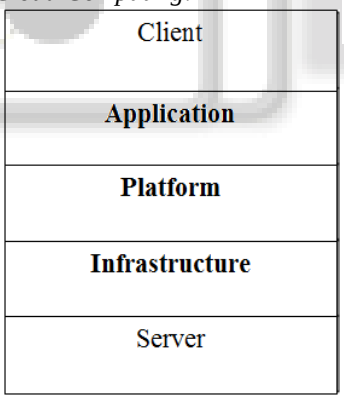

Fig. 1: Layers in Cloud Computing:

- Client – The mobile or desktop system which is used to connect to the cloud services.
- Application – The software application which resides on the cloud.
- Platform – The framework which can be used by cloud applications (Google Apps).
- Infrastructure – The virtual hardware which could act as private cloud servers.
- Server – The physical hardware at the Cloud Service Provider's data center.

7) Major Cloud Service Providers:

Some of the major Cloud Service Providers are:

- Amazon
- Rack space Cloud
- Sales force
- Google

8) Amazon EC2:

- Amazon Elastic Compute Cloud (Amazon EC2) is a web service that provides resizable compute capacity in the cloud. It is designed to make web-scale computing easier for developers.
- Amazon EC2's simple web service interface allows you to obtain and configure capacity with minimal friction.
- Amazon EC2 provides you with complete control of your computing resources and lets you run on Amazon's proven computing environment.
- Amazon EC2 reduces the time required to obtain and boot new server instances to minutes, allowing you to quickly scale capacity, both up and down, as your computing requirements change.
- Amazon EC2 changes the economics of computing by allowing you to pay only for capacity that you actually use. Amazon EC2 provides developers the tools to build failure resilient applications and isolate themselves from common failure scenarios

9) Rack space Cloud:

a) Cloud Servers on Demand:

- Our scalable, affordable, and cloud-driven platform of virtualized servers. Customize and spin up new instances in minutes, or take them down, all with root access, easy-to-use management tools and, of course, our Fanatical Support. You only pay for what's provisioned.

b) Cloud Files:

- Unlimited File Storage & Hosting. Our simple, scalable, and cost-effective online storage solution that leverages the power of the Cloud. Whether your storage needs are modest—or monumental—you enjoy built-in redundancy, an easy-to-use control panel, and Fanatical Support from day one.

c) Cloud Sites:

- Quickly host scalable and reliable websites. Our robust, fast, and easy to use web hosting platform. With built-in redundancy, clustering, and the power of cloud computing, your websites are ready to grow with your business.

10) Service Models:

Once a cloud is established, how its cloud computing services are deployed in terms of business models can differ depending on requirements. The primary service models being deployed are commonly known as:

- Software as a Service (SaaS) — Consumers purchase the ability to access and use an application or service that is hosted in the cloud. A benchmark example of this is Salesforce.com, as discussed previously, where necessary information for the interaction between the consumer and the service is hosted as part of the service in the cloud.
- Platform as a Service (PaaS) — Consumers purchase access to the platforms, enabling them to deploy their own software and applications in the cloud. The operating systems and network access are not managed by the consumer, and there might be constraints as to which applications can be deployed.

- Infrastructure as a Service (IaaS) — Consumers control and manage the systems in terms of the operating systems, applications, storage, and network connectivity, but do not themselves control the cloud infrastructure.

11) *Deployment Models in Cloud Computing:*
- Public Cloud – Cloud Services which are accessible to all for free or payment.
- Private Cloud – Cloud Service accessible only to a single organization.
- Community Cloud – Cloud Services which are accessible to a select few organizations.
- Hybrid Cloud – Private & Public Cloud joins together to provide a single common service.

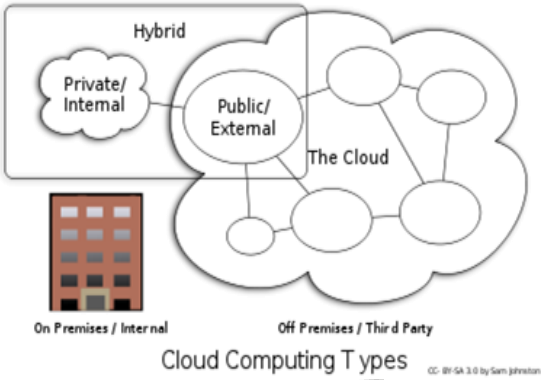

Fig. 2: Cloud Computing Types

12) *Advantages of Cloud Computing:*
- No capital investment needed for technology infrastructure.

13) *Advantages of Cloud Computing:*
- No capital investment needed for technology infrastructure.
- No in-house man power needed to control over the technology infrastructure.
- Need not worry about software upgrades or version changes.
- Pay only for recourses that we use (utility computing).

14) *Disadvantages of Cloud Computing:*
- Privacy: No guaranteed since Cloud Service Provider can monitor your activities.
- Data Security: Cannot guarantee misuse of data at data centers.
- Data Theft: Hacking is on the increase and all data is exposed on the internet.

15) *How Does Cloud Computing Work:*

Fig. 3: Cloud Computing Work

Cloud computer works by hosting your information on computers 'out there' 'in the cloud'. The cloud is basically a batch of computers called data centers or servers that hold your information (files, images, videos, etc) and can be located anywhere. You're working in a cloud because you don't have to store software or files on your own computer. To understand it works, it is important to think of cloud computing in two levels.
- Front level
- Backend

II. RELATED LITERATURE

In literature many techniques have been proposed HSDRT[1], DATA BLOCK RECOVERY[2],ERGOT[3], Linux Box [4], Cold/Hot backup strategy [5],SBBR[6] etc. that, discussed the data recovery process. However, still various successful techniques are lagging behind some critical issues like implementation complexity, low cost, security and time related issues. To cater this issues, in this project we propose a smart remote data backup algorithm, Seed Block Algorithm (SBA). The contribution of the proposed SBA is twofold; first SBA helps the users to collect information from any remote location in the absence of network connectivity and second to recover the files in case of the file deletion or if the cloud gets destroyed due to any reason.

A. High Security Distribution and Rake Technology (HS-DRT):

The HS-DRT [1] is an innovative file back-up concept, which makes use of an effective ultra-widely distributed data transfer mechanism and a high-speed encryption technology, It consists of 3 components: First, the main functions are Data Centre, Second, Supervisory server and third, various client nodes specified by admin. The client nodes are composed of PCs, smart phones, Network Attached Storage and storage service [1]. They are connected with a supervisory server in addition to the Data Centre via a secure network. The basic procedure in the proposed network system is as follows in two sequences one is Backup sequence and second is Recovery sequence. In Backup sequence, when the Data Center receives the data to be backed-up, it encrypts scrambles, divides into some fragmentations, and thereafter duplicates that data to some extents to satisfy with the required recovery rate according to the pre-determined service level. The Data Center encrypts the fragmentations again at the second stage and distributes them to the client nodes in a random order. At the same time, the Data Center sends the metadata used for deciphering the series of fragments. The metadata are composed of encryption keys (both at the first and second stages), several related information of fragmentation, duplication, and distribution [1]. In Recovery Sequence, it is the recovery process when some disasters occur or periodically, the Supervisory Server starts the recovery sequence. It collects the encrypted fragmentations from various appropriate clients like rake reception procedure and they are decrypted, merged, and descrambled in the reverse order at the second stage and the decryption will be completed. Though these processes, the Supervisory Server can recover the original data that should be backed-up. However there are some limitation in this model and

therefore, this model is somehow unable to declare as perfect solution for back-up and recovery. These are: First, in order to fully utilize the HS-DRT processor, the web applications are necessary to be well adjusted to use the HS-DRT engine. Second, is that when the number of duplicated copy of file data increases the corresponding processor performance will be degraded accordingly for executing the web application.

B. Data Block Recovery:

When a data block is corrupted, it can be recovered using the parity block provided by the PCS server and encoded data blocks provided by other nodes in the parity group. Assume that the n-th data block in nodei, Bin, has been corrupted. Nodei sends a recovery request message to the PCS server. On receiving the recovery request message, the PCS server identifies to which VDPG the node belongs to and reads the corresponding parity block, Pn. Then, it generates a temporary random block, r, and a temporary parity block, Pr, for recovery process. When the size of the VDPG is even, Pr = Pn r. Otherwise, Pr = Pn. The PCS server sends Pr along with the list of nodes that will send their encoded data block to nodei for recovery along with the IP address of nodei to all other nodes in the group. If there are any off-line nodes, the PCS server sends the message when they become on-line. On receiving the message, each node generates their own encoded data block, E, by XORing the n-the data block with r (Ej = Bin r, for each node j VDPG, j i) and sends to nodei. Then, the nodei recovers the corrupted data block by Bin = PrE1⋯ Ei-1 Ei+1⋯ E |VDPG|. (1) Note that the whole virtual disk corruption can be recovered by iterating the above data block recovery process.

C. Efficient Routing Grounded on Taxonomy (ERGOT):

Efficient Routing Grounded on Taxonomy [4] is a Semantic-based System for Service Discovery in Distributed Infrastructures in cloud computing. In our survey, we found a unique way of data retrieval. We made a focus on this technique as it is not a back-up technique but it provide an efficient retrieval of data that is completely based on the semantic similarity between service descriptions and service requests. It also exploits both coarse-grain service functionality descriptions and at a finer level. ERGOT is built upon 3 components. These components include: 1) A DHT (Distributed Hash Table) protocol, which we use to advertise semantic service description annotated using concepts from ontology, 2) A SON (Semantic Overlay Network), enables the clustering of peer that have semantically similar service description. The SON is constructed incrementally, as a product of service advertising via DHT, 3) A measure of semantic similarity among service description [4]. DHTs and SONs both networks architectures have some shortcomings. Hence, ERGOT combines both these network Concept. The ERGOT system proposed semantic-driven query answering in DHT-based systems by building a SON over a DHT. An extensive evaluation of the system in different network scenarios demonstrated its efficiency both in terms of accuracy of search and network traffic. DHT-based systems perform exact-match searches with logarithmic performance bounds, however does not go well with semantic similarity search models.

D. Linux Box:

Another technique to reduces the cost of the solution and protect data from disaster. It also makes the process of migration from one cloud service provider to other very easy. It is affordable to all consumers and Small and Medium Business (SMB). This solution eliminates consumer's dependency on the ISP and its associated backup cost. A simple hardware box can do all these at little cost named as simple Linux box which will sync up the data at block/file level from the cloud service provider to the consumer. It incorporates an application on Linux box that will perform backup of the cloud onto local drives. The application will interface with cloud on a secured channel, check for updates and sync them with local storage. The data transmission will be secure and encrypted. After a valid login, the application secures the channel using IP Security and in-flight encryption techniques. The application then interacts with the application stack at the cloud service provider and does a onetime full backup. During subsequent check, it backs up only the incremental data to the local site. The limitation we found that a consumer can backup not only the Data but Sync the entire Virtual Machine[5] which somehow waste the bandwidth because every time when backup takes place it will do back-up of entire virtual machine.

E. Cold and Hot Backup Service Replacement Strategy (CBSRS):

In Cold Backup Service Replacement Strategy (CBSRS) recovery process, it is triggered upon the detection of the service failures and it will not be triggered when the service is available. In Hot Backup Service Replacement Strategy (HBSRS), a transcendental recovery strategy for service composition in dynamic network is applied [6]. According to the availability and the current state of service composition before the services interrupt, it restores the service composition dynamically. During the implementation of service, the backup services always remain in the activated states, and then the first returned results of services will be adopted to ensure the successful implementation of service composition. On Comparing HBSRS with the CBSRS, it reduced service recovery time. However, because backup services and original services are executed at the same time, the recovery cost increases accordingly.

F. Shared Backup Router Resources (SBBR):

In one of our survey, we found that one technique basically focuses on the significant cost reduction and router failure scenario i.e. (SBBR). It concerns IP logical connectivity that remains unchanged even after a router failure and the most important factor it provides the network management system via multi-layer signaling .However it concerns with the cost reduction concept there exist some inconsistencies between logical and physical configurations that may lead to some performance problem. Additionally [10], it show how service imposed maximum outage requirements that have a direct effect on the setting of the SBRR architecture (e.g. imposing a minimum number of network-wide shared router resources locations).However, it is unable to includes optimization concept with cost reduction

III. REMOTE DATA BACKUP SERVER

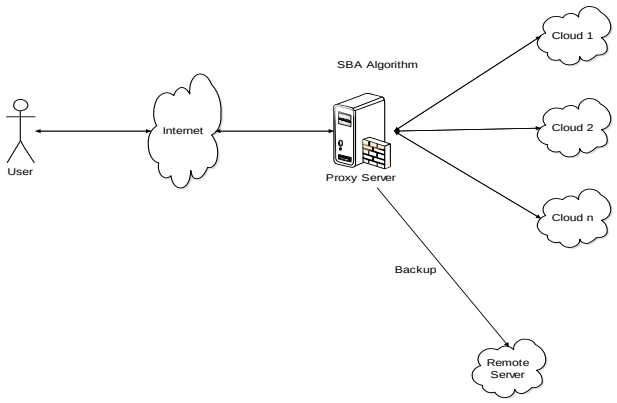

Fig. 4: Remote data Backup Server and its Architecture

The remote backup services should cover the following issues

A. Data Integrity:

Data Integrity is concerned with complete state and the whole structure of the server. It verifies that data such that it remains unaltered during transmission and reception.it is the measure of the validity and fidelity of the data present in the server.

B. Data Confidentiality:

Sometimes client's data files should be kept confidential such that if no.of users simultaneously accessing the cloud, then data files that are personal to only particular client must be able to hide from other clients on the cloud during accessing of file.

C. Trustworthiness:

The remote cloud must possess the Trustworthiness characteristic. Because the user/client stores their private data; therefore the cloud and remote backup cloud must play a trustworthy role.

D. Cost efficiency:

The cost of process of data recovery should be efficient so that maximum no. of company/clients can take advantage of back-up and recovery service. There are many techniques that have focused on these issues. In forthcoming section, we will be discussing a technique of back-up and recovery in cloud computing domain that will cover the for said issues.

IV. DESIGN OF THE PROPOSED SEED BLOCK ALGORITHM

- Initialization: Main cloud: Mc;
- Remote Server: Rs;
- Client of Main Cloud: Ci;
- Files:a1 and a1';
- Seed block: si;
- Random Number: r;
- Client's ID: client_Idi
- Input: a1 created by Ci; r is generated at Mc;
- Output: Recovered file a1after deletion at Mc
- Given: Authentication clients could allow uploading, downloading and do modification on its own the files only.

1) Step 1: Generating a random number
   int r=rand();
2) Step 2: Create a Seed Block Si for each Ci and store
   Si at Rs.
   Si=r Exor Client_Idi(Repeat step 2 for all clients)
3) Step 3: if Ci/Admin creates/modify files a a1 and stores
   Mc,then a1' create as,
   a1'=a1 EXOR Si
4) Step 4: Store a1' at Rs.
5) Step 5: If server crashes a1 deleted from Mc, then,We do EXOR to retrieve the original a1 as:
   a1=a1' EXOR Si
6) Step 6: Return a1 to Ci
7) Step 7: END

A. Module 1:

1) Cloud Interface Creation:

The client wants to create a file in the cloud. Cloud requires the Id of the client and a random number to generate the seed block for respective clients. Cloud generates a random number for the corresponding client and also produces the seed block which is obtained by EXORing the client Id with the random number that is generated.

B. Module 2:

1) Seed Block Algorithm Initialization:

The client wants to create a file in the cloud. Cloud requires the Id of the client and a random number to generate the seed block for respective clients. Cloud generates a random number for the corresponding client and also produces the seed block which is obtained by EXORing the client Id with the random number that is generated.

C. Module 3:

1) File Upload:

When the client wants to upload a file in the cloud the client provides its Id along with the file to be uploaded. Cloud performs the EXOR operation in which the client's file is EXORed with the respective seed block of the client. The EXORed file is stored.

D. Module 4:

1) File Download:

When the client wants to download a file the client provides its Id and the file name to the cloud. The cloud in turn finds the appropriate seed block of the client and performs the EXOR operation of the seed block with the file that is stored in the cloud and which is requested by the client. After performing the EXOR operation the file in the cloud is checked to see whether it matches the file specified by the user. The file is provided to the client.

E. Module 5:

1) Storage Node Repair:

When the client wants to download a file the client provides its Id and the file name to the cloud. The cloud in turn finds the appropriate seed block of the client and performs the EXOR operation of the seed block with the file that is stored in the cloud and which is requested by the client. After performing the EXOR operation the file in the cloud is checked to see whether it matches the file specified by the user. The file is provided to the client.

V. CONCLUSION AND RESULT

In this project, we presented detail design of proposed SBA algorithm. Proposed SBA is robust in helping the users to collect information from any remote location in the absence of network connectivity and also to recover the files in case of the file deletion or if the cloud gets destroyed due to any reason. Experimentation and result analysis shows that proposed SBA also focuses on the security concept for the back-up files stored at remote server, without using any of the existing encryption techniques. The time related issues are also being solved by proposed SBA such that it will take minimum time for the recovery process. The experiment results show that the file stored in the cloud is retrieved from the backup server without any redundancy in an efficient manner.

REFERENCE

[1] Henry C. H. Chen and Patrick P. C. Lee Department of Computer Science and Engineering, the Chinese University of Hong Kong ":Enabling Data Integrity Protection in Regenerating-Coding-Based Cloud Storage"[1]
[2] Kevin M. GreenanParaScale, Inc.James S. Plank University of TennesseeJay J. WylieHP Labs "Mean time to meaningless: MTTDL, Markov models, and storage system reliability" [2]
[3] Kruti Sharma, Kavita R Singh Computer Science Engineering, YCCE, Nagpur (M.S), India" Online Data Back-up and Disaster Recovery Techniques in Cloud Computing: A Review"
[4] Yuchong Hu, Henry C. H. Chen, Patrick P. C. Lee, Yang Tang the Chinese University of Hong Kong, Columbia University "NC Cloud: Applying Network Coding for the Storage Repair in a Cloud-of-Clouds" [3]

