ISSN 2278-3091

International Journal of Advanced Trends in Computer Science and Engineering, Vol. 3 , No.1, Pages : 187– 192 (2014)
Special Issue of ICETETS 2014 - Held on 24-25 February, 2014 in Malla Reddy Institute of Engineering and Technology, Secunderabad– 14, AP, India

Provide Security and Data Backup For Cloud Computing using RDBS

Mr.M.Ganesh kumar[1] Mr.D.Kalyan Kumar [2] Mr.K.Ramesh Babu[3]

1:Asst.professor in Malla Reddy Institute Of Engineering And Technology,ganeshkumar.programs@gmail.com.
2: Asst.professor in Malla Reddy Institute Of Engineering And Technology,dkkumar123@gmail.com.
3. Asst.professor in Malla Reddy Institute Of Technology And Science, rameshbabu.kanaparthi@gmail.com.

Abstract --- Today in cloud computing, generate the data in e-form are huge in amount. to sustain this data powerfully, there is an essential of data recovery services. To cater this, in this article we recommend a remote data back up server(RDBS). The remote data back up server is a server which holds the main cloud's entire data as a whole and locate at remote place(remote from cloud). And if the central repository lost its data, then it uses the information from the remote repository. The purpose is to help clients to gather information from remote storage area either if network connectivity is not existing or the main cloud is unable to make available the data to the clients in RDBS, if clients found that data is not available on central repository, then clients are allowed to access the files from remote repository (i.e. indirectly). The time related issues also being solved by an RDBS such that it will take minimum time for the recovery process. Recommended RDBS also focuses on the security concept for the backup files stored at remote server, without using of the existing encryption techniques

Keywords: privacy; central repository central repository; remote repository, parity cloud, parity cloud service, RDBS.

1. INTRODUCTION

National Institute of Standard and Technology defines as a model for enabling convenient, on-demand network access to a share pool of configurable computing service (for example, networks, servers, storage, applications and services) that can be provisioned rapidly and released with minimal management effort or services provider [1]. Cloud computing is no longer a buzzword today. In addition, it changes and improves the way we uses the computing platform.

In today's world, there is a huge increase in the electronic data. This requires large huge of data storage devices to store this huge amount of data. These requirement leads to introduction of 3 Tera Byte HDD. Therefore, usually consumer prefers to store large amount of private data in cloud. Unfortunately, if cloud will be corrupted or damaged it leads to the loss of all important and private data then there should be some mechanisms to take back-up of the data, and provide the data at the time of cloud failure or loss of data.

As we know that plain data back-up techniques are having many reliability and security problems. However the plain back-up techniques are not convenient and reliable as well. So to overcome from plain data backup and recovery problem, it requires more safe and effective system such as RAID (Redundant Array Independent Disk), HSDRT [1], PCS [2], ERGOT [3], Linux Box [5], Cold and Hot back-up technique [6], SBBR [10], REN [17] etc. These systems provide high privacy protection and reliability however some increases the cost where as some are unable to maintain the implementation complexity low.

Although many backup and recovery techniques have been proposed during last few years in the computing domain; however, real world scenarios remain a challenge. In this review paper, we focuses on the various techniques of back-up and recovery on cloud computing. Each technique is greatly affected in real time scenario either in redundancy point of view or security point of view or complex algorithm's implementation point of view. This paper is organized as follows: Section II explains the need of cloud computing. In Section III we discuss about the Remote Data Backup Server. The existing methods that are successful to some extent in the cloud computing domain are reviewed in Section IV. Finally, in Section V discussion and conclusions are given.

ISSN 2278-3091
International Journal of Advanced Trends in Computer Science and Engineering, Vol. 3 , No.1, Pages : 187– 192 (2014)
Special Issue of ICETETS 2014 - Held on 24-25 February, 2014 in Malla Reddy Institute of Engineering and Technology, Secunderabad– 14, AP, India

II. RELATED LITERATURE

In literature, we study most of the recent back-up and recovery techniques that have been developed in cloud computing domain such as HSDRT[1], PCS[2], ERGOT[4], Linux Box [5], Cold/Hot backup strategy [6] etc. Detail review shows that none of these techniques are able to provide best performances under all uncontrolled circumstances such as cost, security, low implementation complexity, redundancy and recovery in short span of time.

Among all the techniques reviewed PCS is comparatively reliable, simple, easy to use and more convenient for data recovery totally based on parity recovery service. It can recover data with very high probability. For data recovery, it generates a virtual disk in user system for data backup, make parity groups across virtual disk, and store parity data of parity group in cloud. It uses the Exclusive–OR ( ) for creating Parity information. However, it is unable to control the implementation complexities.

On the contrary, HSDRT has come out an efficient technique for the movable clients such as laptop, smart phones etc. nevertheless it fails to manage the low cost for the implementation of recovery and also unable to control the data duplication. It an innovative file back-up concept, which makes use of an effective ultra-widely distributed data transfer mechanism and a high-speed encryption technology 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. This proposed system follows two sequences one is Backup sequence and second is Recovery sequence. In Backup sequence, it receives the data to be backed-up and in Recovery Sequence, when some disasters occurs or

periodically, the Supervisory Server (one of the components of the HSDRT) starts the recovery sequence. 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.

Rather, Efficient Routing Grounded on Taxonomy (ERGOT) [4] is totally based on the semantic analysis and unable to focus on time and implementation complexity. It is a Semantic-based System which helps for Service Discovery in cloud computing. Similarly, 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.

ERGOT is built upon 3 components 1) A DHT (Distributed Hash Table) protocol 2) A SON (Semantic Overlay Network), 3) A measure of semantic similarity among service description [4]. Hence, ERGOT combines both these network Concept. By building a SON over a DHT, ERGOT proposed semantic-driven query answering in DHT-based systems. However does not go well with semantic similarity search models.

In addition, Linux Box model is having very simple concept of data back-up and recovery with very low cost. However, in this model protection level is very low. 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. It 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 data transmission will be secure and encrypted.

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. Similarly, we also found that one technique basically focuses on the significant cost reduction and router failure scenario i.e. (SBBR). It concerns IP logical connectivity that will be remain unchanged even after a router failure and the most important factor is that it provides the network management system via multi-layer signaling.

II. NEED FOR BACK-UP IN CLOUD COMPUTING

Cloud computing provides on demand resources to the consumer/user. It requires the management of resources among each and every client/user. Such management includes various aspects of proper utilization of the resources. The resources can be any hardware or software. The software like any application programming interface, application development kit and any type of data file etc. Various choices are there among various implementations for back up of the data and that maintain its security among various users. Cloud computing must be able to provide reliability such that users can upload their sensitive and important data. The cost-effective approach is the main concern while implementing any cloud.

During the study of cloud computing, we found various advantages of cloud computing. In advantages, we found that the cloud is capable enough to store the huge amount of data of various different clients with complete security such that Internet Service Provider (ISP) provides a huge storage in a cloud to the user. And users are allow to

ISSN 2278-3091
International Journal of Advanced Trends in Computer Science and Engineering, Vol. 3 , No.1, Pages : 187– 192 (2014)
Special Issue of ICETETS 2014 - Held on 24-25 February, 2014 in Malla Reddy Institute of Engineering and Technology, Secunderabad– 14, AP, India

upload there private and important data to the main cloud. And at the same time we found critical issue regarding this storage i.e. if any of the client's data file is missing or disappeared for some reason or the cloud get destroyed either due to any natural calamity (like flood, earthquake etc.), then for back-up and recovery consumer/client has to depend on service provider which means the data has to be stored in the server.

To overcome problem of such scenario, it requires an efficient technique for data backup and recovery so that the client can able to contact the backup server where private data is stored with high reliability and whenever a main cloud fails to provide the user's data. These techniques must possess the

low cost as well for implementation of the recovery problem's solution and can easily recover the data after any disaster. That's why, the need of the back-up and recovery techniques for cloud computing arises due to heavy storage of its clients.

III. REMOTE DATA BACK-UP SERVER

Remote Data Backup server is a server which stores the main cloud's entire data as a whole and located at remote place (far away from cloud). And if the central repository lost its data, then it uses the information from the remote repository. The purpose is to help clients to collect information from remote repository either if network connectivity is not available or the main cloud is unable to provide the data to the clients. As shown in Fig 1, if clients found that data is not available on central repository, then clients are allowed to access the files from remote repository (i.e. indirectly).

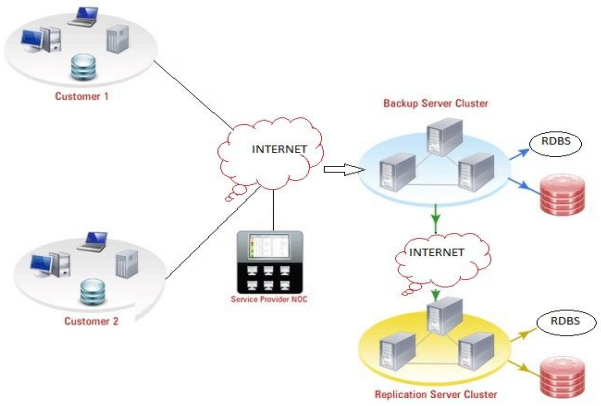

Fig. 1. RDBS Architecture

EXISTING METHODS

In our research, we find out many techniques that are having their different customs to create retrieving and recovery backup. Finally speaking, all those techniques shows on three different aspects, such as security & privacy issues cost controlling, data duplication. All the technique has the complete focal point on their aim of backup and recovery. Further, we detail few recent techniques PCS [1], ERGOT [2], Linux Box [3], Cold and Hot back-up technique [6], SBBR [10], REN [17] that have address the abovementioned issues.

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

The HS-DRT [1] is an inventive file back-up concept, which make use of an effective ultra-widely distributed data transport 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 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

ISSN 2278-3091
International Journal of Advanced Trends in Computer Science and Engineering, Vol. 3 , No.1, Pages : 187– 192 (2014)
Special Issue of ICETETS 2014 - Held on 24-25 February, 2014 in Malla Reddy Institute of Engineering and Technology, Secunderabad– 14, AP, India

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. Parity Cloud Service Technique

Parity Cloud Service technique (PCS) [2] is a very simple, easy to use and more convenient for data recovery which is based on parity recovery service. A PCS has low cost for recovery and can recover data with very high probability. For data recovery, PCS uses a new technique of generating virtual disk in user system for data backup, make parity groups across virtual disk, and store parity data of parity group in cloud. The algorithms for PCS work as follows by using the Exclusive–OR () for creating Parity information.

1) Initial Parity Generation:

In this, the seed block (Si) is generated for virtual disk. PCS server sends the initialize message to each Recovery Manager in the group. After sending the initialize message, the PCS server sends temporary random block (r) to the first node. On receiving the r block, the 1st node (node 1) generates an intermediate parity block via r S1 and sends it to its successor, node 2. Accordingly, node 2 generates an intermediate parity block via XORing the received parity block with its seed block, S2, and sends it to its successor, node 3 and so on. The final block transferred to the PCS server from node 4 is XORed with the temporary random block, r, again, to generate the seed parity block across all
seed blocks (((((r S1)S2) S3) S4) r = S1S2 S3S4). The initialization process occurs only once for each parity group. The seed parity block stored separately from in the metadata region of each virtual disk, for later use.

2) Parity Block Update:

The Storage Manager in PCS agent maintains parity generation bitmap (PG-bitmap) .It indicate whether the parity block for each data block in the virtual disk has been generated or not. The bitmap is initialized (set to 0) after the initialization process for any data block in the virtual disk. The PG-bitmap is referred when a block is updated. When a block (Bold) in nodei is to be updated to a new block (Bnew), the Storage Manager refers to the corresponding value in the PG-bitmap. If it is 0, then the Storage Manager generates an intermediate parity block (Pt) by XORing the new block with the seed block (Pt = Bnew Si), and set the corresponding value in the PG-bitmap to 1. Otherwise, the intermediate parity block is generated by XORing the new block and the old block (Pt = Bnew _ Bold). For each VDPG, the PCS server also maintains the PG-bitmap. Note that the parity block update can be easily done by the data updating node and the PCS server. In the updating process, other nodes are not needed to participate.

3) 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.

Apart from its best performance given by the algorithm discussed above PCS somehow lags behind in providing perfect solutions to backup and recovery due to some limitations. These limitations are [2]: first one is that, the recovery process cannot finish if one or more participating nodes are not online at the recovery time. Second limitation is that, PCS is based on Markov process and calculate Mean according to the group size. Generally, it fails to precisely estimate the real data reliability of PCS.

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

ISSN 2278-3091
International Journal of Advanced Trends in Computer Science and Engineering, Vol. 3 , No.1, Pages : 187– 192 (2014)
Special Issue of ICETETS 2014 - Held on 24-25 February, 2014 in Malla Reddy Institute of Engineering and Technology, Secunderabad– 14, AP, India

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.

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

G. Rent out the Rented Resources

Another technique we found in the field of the data backup is a REN (Research Education Network) cloud. As we know the Cloud services are expensive and large number of enterprises and individuals are attracted towards low cost cloud services. The lowest cost point of view we found a model "Rent out the Rented Resources" [17]. It aims to reduce the monetary cost of cloud services. They have proposed a three phase model for cross cloud federation. These three phases are discovery, matchmaking and authentication. Keahey et. al. introduced the concept of Sky Computing [15].This model is based on concept of cloud vendor that rent the resources from venture(s) and after virtualization, rents it to the clients in form of cloud services. The cooperating venture is paid for its infrastructure utilization [17]. It is based on three core objectives: 1) It minimizes the cloud infrastructure cost. 2) It provides low cost cloud services by reducing infrastructure cost for the cloud vendors to the clients. 3) It gives the monetary benefit with the large under-utilized technology infrastructure to the established enterprises (cooperating ventures).

ISSN 2278-3091

International Journal of Advanced Trends in Computer Science and Engineering, Vol. 3 , No.1, Pages : 187– 192 (2014)
Special Issue of ICETETS 2014 - Held on 24-25 February, 2014 in Malla Reddy Institute of Engineering and Technology, Secunderabad– 14, AP, India

V. DISCUSSION AND CONCLUSION

In this paper all these approach are give good performance low completion complexity under all uncontrolled situation such as charge security, redundancy and revival in short span of time. Among all the techniques maintain its privacy for each resource and also it tries to minimize the cost of infrastructure. However, it is unable to control the implementation complexities. On the contrary, HSDRT has come out an efficient technique for the movable clients such as laptop, smart phones etc. nevertheless it fails to manage the low cost for the implementation of the recovery and also unable to control the data duplication. Rather, ERGOT is totally based on the semantic analysis and unable to focus on time and implementation complexity. In addition, Linux Box model is having very simple concept of data back-up and recovery with very low cost. However, in this model protection level is very low. Similarly, in the list of techniques maintaining the cost of implementation, SBBR focuses on the cost reduction; however fails to concentrate on the optimization concept and redundancy. With entirely new concept of virtualization REN cloud also focuses on the low cost infrastructure with the complex implementation and low security level. All these techniques tried to cover different issues maintaining the cost of implementation as low as possible. However there are some techniques in which cost increases gradually as data increases.

REFERENCES

[1] Yoichiro Ueno, Noriharu Miyaho, Shuichi Suzuki,Muzai Gakuendai,
Inzai-shi, Chiba,Kazuo Ichihara, 2010, “Performance Evaluation of a
Disaster Recovery System and Practical Network System
Applications,” Fifth International Conference on Systems and
Networks Communications,pp 256-259.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] Lili Sun, Jianwei An, Yang Yang, Ming Zeng, 2011, “Recovery
Strategies for Service Composition in Dynamic Network,”
International Conference on Cloud and Service Computing.