JCSE International Journal of Computer Sciences and Engineering Open Access

Survey Paper Vol.-6, Issue-5, May 2018 E-ISSN: 2347-2693

A Survey on Data Recovery Approaches in Cloud Computing Environment

I. Benjamin Franklin [1*], T.N. Ravi [2]

[1]Manonmaniam Sundaranar University, Tirunelveli
[1] Department of Computer Applications, St. Joseph's College of Arts & Science (Autonomous), Cuddalore
[2] PG and Research Department of Computer Science, Periyar EVR College (Autonomous), Trichy

[*]Corresponding Author: franklinbenj@gmail.com

Available online at: www.ijcseonline.org

18/May/2018, Published: 31/May/2018

Abstract—Cloud computing system provides a lot of convenient computational and data storage services to the users. Data transfer to the cloud environment is convenient. The cloud computing system generates a large amount of private data on the main cloud. Then, the need of data recovery services are increasing day-by-day and require development of efficient data recovery technique. The aim of the data recovery technique is to collect the information from the backup server, when the server lost the data and unable to provide the data to the user. Various techniques are proposed for efficient recovery of data. This paper focuses on the comprehensive review of the data recovery approaches, issues in data recovery, failures in cloud storage, key factors and their role in data recovery and existing data security technologies in the cloud. The main objective of the review paper is to summarize the prevailing data recovery techniques in the cloud computing domain.

Keywords— Cloud Computing, Cloud Storage, Data Recovery, Data Security, Data Storage Service, Private Data

I. INTRODUCTION

Cloud computing provides dynamic and scalable virtualization resources and services to the users in a pay-per-use basis through Internet. The characteristics of the cloud computing environment include integrated server clusters, on-demand services, network dependency, resource virtualization, centralized computing and shared server with multi-tenancy based on the definition and service models. Cloud storage is a new concept extended from the domain of cloud computing system [1]. The cloud storage is a network equipment, storage device, servers, applications, public access interfaces and client software. A large number of storage devices are configured in the cloud computing system while storing and managing a huge amount of data. Cloud storage stores the local data in the online storage space provided by the Storage Service Provider (SSP) through network. Users do not need to construct their own data centres. The cloud storage avoids the duplication of storage platforms and saves the software and hardware infrastructure cost. To provide data storage services, the cloud storage facilitates collaboration between different types of storage devices. Compared to the traditional storage methods, the cloud storage poses new challenges in data security, reliability and management. The cloud storage uses the application software for the collaboration of different storage devices and provision of data storage functions. To ensure data security and business continuity, there is a need to build the data backup systems and recovery systems. All the devices are completely transparent to the user in the cloud storage system. Any authorized user can connect with cloud storage through a cable and access data on the cloud storage in any location [2].

With the rapid progress of the information society and increase in the number of global operations, a huge business data has become the core of global business connections and data storage security becomes particularly important. When the data that is not used temporarily but significant, it will be stored in the cloud storage. The cloud stores the user data using distributed file system with a large number of disks. The users have to face the issues such as unreliable data storage, disclosure of personal privacy and exposure to security threats. Thus, there arises the need for developing physical, logical, and personnel access control policies. Data security issues arise in the cloud storage due to the lack of authentication, audit control, feeble encryption algorithms and cryptographic keys, risk of association, unreliable data center and lack of disaster recovery [3]. The data protection, integrity, confidentiality, data breaches and data recovery approaches are applied for addressing these issues [4].

Various solutions are proposed for the data privacy and network attacks [5-9]. With the increase in the growth of data, there is a vital need to ensure the prevention of data loss.

Sarang and Bunkar [10] presented a survey about the service models and privacy issues in the cloud computing environment. The challenges and opportunities in the cloud environment are also identified. Kadam and Kumbhkar [11] provided a brief introduction of the types and security issues to secure the data in the cloud environment. The simulation of the Distributed Denial of Service (DDoS) attack using the MATLAB software is done to find the quantitative measure of its effect on the victim [12].

This paper mainly focuses on the related data recovery approaches that ensure continuation of services even during the loss of data. The data backup and recovery protocols and the technology of isolating stored data and copies need to be studied. In this paper, we focus on the failures in cloud storage, factors on data recovery and present a survey of existing solutions for data recovery.

The paper is organized as: Section II describes a brief overview of the current data recovery techniques and the factors in data recovery are mentioned in Section III. The issues in data recovery are mentioned in Section IV. Section V states the failures in cloud storage. The conclusion of the review is discussed in Section VI.

II. Literature Review

A. *Seed Block Algorithm (SBA)*

SBA [13-17] provides the backup and recovery process using the Exclusive OR (XOR) operation. The cloud architecture comprises main cloud, clients and remote server. Here, a random number and unique client ID are set in the main cloud and client, respectively. While storing the client ID, XOR operation is performed on the client ID and random number to create a seed block for that particular client. The generated seed block is stored at the remote cloud server. The client creates a file for the first time and stores the file at the main cloud. XOR operation is performed at the main file with the seed block and stored at the remote cloud server, when the file is stored in the main cloud server. If the file in the main cloud is damaged or deleted erroneously, then the user will obtain the original file by performing XOR of the file with the seed block of the corresponding client to create the original file and return the resulted file i.e. original file back to the requested client.

SBA Algorithm

```
Initialization: Main cloud $M_c$, Remote Server $R_S$, clients $C_i$, Files $a_1$ and $a_1'$, seed block $S_i$, random number ‘r’ and client ID $Client\_Id_i$
Input: $a_1$ created by $C_i$; r is generated at $M_c$
Output: Recovered file $a_1$ after deletion of file at the main cloud
Authenticated files could allow uploading, downloading and perform self-modification of the files.
Step 1: Generate a random number $int\ r = rand(\,);$
Step 2: Generate $S_i$ for each client and store the $S_i$ at the remote cloud server
$S_i = r \oplus Client\_Id_i$
Step 3: If $C_i/Admin$ creates or modifies the files and stores at the main cloud, then $a_1'$ is created as
$a_1' = a_1 \oplus S_i$
Step 4: Store $a'$ at the remote server
Step 5: If the server crashes the files deleted from the main cloud, then EXOR is applied to retrieve the original file as
$a_1 = a_1' \oplus S_i$
Step 6: Return the file $a_1$ to the client
Step 7: End
```

B. *Parity cloud service (PCS)*

Song et al. [18] devised a novel data recovery framework for the cloud environment. It generates virtual disk belonging to the Virtual Disk Parity Group (VDPG) for the private backup, creates a parity group through the virtual disks, and stores the parity data in the storage. The proposed framework is simple and does not consume more resources for privacy protection. It does not require uploading of data to the server for data recovery. The data loss rate is reduced by using the collaboration-based data recovery algorithm [19].

When a data block is corrupted, the block can be recovered using the parity block and encoded data blocks provided by other nodes in the parity group. Figure.1 shows the process for data block recovery in the PCS. Let us assume that the data block ‘n’ in the node ‘i’, B_n^i is corrupted. Then, the node ‘i’ sends a recovery request message containing the corrupted block number to the PCS server. While receiving this message, the PCS server identifies which VDPG the node belongs to and reads the corresponding parity block P_n across all data blocks with the same block number in the virtual disks. Then, a temporary random block ‘r’ is generated for encoding data blocks in other nodes than the node ‘i’. The PCS server generates a temporary parity block P_r as follows

$$P_r = \begin{cases} P_n \oplus r \; if |VDPG| \; is \; even \\ P_n \qquad\qquad Otherwise \end{cases} \qquad (1)$$

The PCS server sends the temporary parity block along with the list of nodes in the parity group to the node ‘i’ for data recovery. While receiving the list, the node ‘i’ waits for the encoded data blocks from other nodes in the list. Then, the PCS server sends a message containing *r*, *n*, and the IP address of the node ‘*i*', to all nodes in the group other than the node *i*. If there are any offline nodes, the PCS server sends the message when the nodes are online. While receiving the message, each node generates own encoded data block using the exclusive OR (XOR) on n with r. For example, the node 'j' generates encoded data block as $E_j = B_n^j \oplus r$.

Each node sends its encoded data block to 'i'. Then, the node 'i' recovers the corrupted data block through the XOR of the temporary parity block and all encoded data blocks received from other nodes

$$B_n^i = P_r \oplus E_1 \oplus \ldots \oplus E_{i-1} \oplus E_{i+1} \oplus \ldots \oplus E_{|VDPG|} \quad (2)$$

P_r always makes 'r' even so that they can be cancelled in the equation. It is noted that since node 'i' does not have 'r', original data blocks of other nodes cannot be extracted from the encoded data blocks. This enables privacy protected data recovery in PCS. The entire virtual disk corruption can be recovered through the iteration of the above data block recovery process. The recovery process cannot be finished if one or mode nodes in the parity group are not online at the data recovery time. Hence, the process finishes as soon as the last offline node becomes online.

C. *Bloom filter*

The bloom filter [20, 21] interacts well with erasure correcting codes. Specifically, the error block is verified when the bloom filter is used. The specific corrupted data blocks are repaired using the recovery method. Initially, the integrity verification is applied for finding the corrupted block. Then, the corrupted block is repaired. The data integrity is checked again after the data recovery to determine whether the repair process is effective or not. The data recovery is related to the integrity verification.

A Third Party Auditor (TPA) generates a repairing set of data blocks $G_1 = \{b_i | n_1 \le i \le n_s\}$ comprising 's' corrupted blocks. Then, the TPA outsources the data blocks G_1 and MapReduce program for repairing the corrupted blocks and updating the bloom filter and Dynamic Storage Table (DST) to the CSP by Remote Procedure call (RPC). The CSP executes the MapReduce programs and outputs the result to a set G_2. G_2 contains the updated value of the filter and DST. After receiving G_2 from the CSP, TPA can easily determine the integrity of data file by verifying whether each signature satisfies the query in counting the filter. The users do not have to check and recover the data files by themselves as the data recovery is set to a periodical task. The integrity of the data file is checked periodically to avoid loss from the false positive in the bloom filter.

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

HS-DRT [22, 23] is an innovative backup concept used for distributed data transfer and high-speed encryption. In the backup sequence, the data is encrypted, divided into fragmentations and duplicated to satisfy the data recovery rate according to the predetermined service level. The data centre encrypts the fragments and distributes them to the client nodes in a random way. It sends the metadata used for decoding the series of encrypted fragments. The supervisory server starts the data recovery sequence, during the occurrence of disasters. The encrypted fragmentations are collected from various clients, decrypted and descrambled in the reverse order and decryption is completed. Though these processes, the supervisory server can recover the original data that should be backed-up.

There are some limitations in this HS-DRT model to act as perfect backup and recovery solution. To completely utilize the HS-DRT processor, the web applications are required to be well adjusted to use the HS-DRT engine. When there is an increase in the number of duplicated data file, the performance of the process is degraded accordingly for executing the web application.

E. *Efficient Routing Grounded On Taxonomy(ERGOT)*

ERGOT system [24-26] combined Semantic Overlay Networks (SONs) and Distributed Hash Tables (DHTs) for the semantic-based service discovery in the distributed infrastructures such as data grids and clouds. The services are advertised in the DHT based on their annotations to establish a SON among the service providers. The semantic-based service matchmaking is enabled using a new similarity measure between the service requests and descriptions. The ERGOT system recommended semantic-driven query answering in DHT-based systems by creating a SON over a DHT. The system achieved high search accuracy and network traffic in different network scenarios. The DHT-based systems perform exact-match searches with the logarithmic performance bounds. However, this does not go well with the semantic similarity search models.

F. *Linux Box*

Linux box [27-29] reduces the cost of the solution and protects data from disaster. It facilitates the migration between the service providers and affordable to all consumers. This eliminates the dependency on the Internet Service Provider (ISP) and associated backup cost. The Linux box synchronizes up the data at the file level from the CSP to the consumer. An application is incorporated on the Linux box to perform backup of the cloud onto the local drives.

The application will interface with the cloud on a secured channel, check for the updates and synchronize the updates with local storage. The data transmission will be secure and encrypted. After a valid login, the application secures the channel using IP Security and onboard encryption techniques. Then, the application interacts with the application stack at the CSP and does a previous complete backup. During successive check, it backs up only the incremental data to the local site. The main limitation is the wastage of the bandwidth during the backup of entire virtual machine.

G. *Cold and Hot Backup Service Replacement Strategy (CBSRS and HBSRS)*

In CBSRS [17, 30-34], the recovery process is triggered during the detection of the service failures. In HBSRS, an inspiring recovery strategy for service composition in the dynamic network is applied. The service composition is restored dynamically according to the data availability and the current state of service composition before the interruption of services. During the implementation of service, the backup services always remain in the activated state. Then, the first returned results of services will be adopted to ensure the successful implementation of service composition. While comparing HBSRS with the CBSRS, it reduced service recovery time. However, because the backup services and original services are executed simultaneously, the recovery cost increases consequently.

H. *Shared Backup Router Resources (SBRR)*

SBRR [35-38] focuses on the significant cost reduction and router failure scenario. It concerns Internet Protocol (IP) logical connectivity that remains unchanged even after a router failure. The most important factor is it provides the network management system through the multi-layer signalling. But, it concerns with the cost reduction concept. The variations between the logical and physical configurations may lead to some performance problem. Furthermore, the direct effect of the service imposed maximum outage requirements on the SBRR architecture. However, it is unable to include optimization concept with cost reduction.

I. *Rent Out the Rented Resources*

As the cloud services are expensive, a large number of individuals are attracted towards the low-cost cloud services. Rent out the Rented Resources [39, 40] aims to reduce the cost of cloud services. It proposed a three phase model including discovery, matchmaking and authentication for cross cloud group. Remus provides extremely high fault tolerance, so that the running system can continue execution on an alternate physical host during the occurrence of failure within only seconds of server downtime, while completely preserving the active network connections. Keahey et al. [41] introduced the sky computing concept based on the concept of renting the resources to the clients in the form of cloud services. This infrastructure is independent of any particular CSP and can be instantiated dynamically. It is based on three objectives

- It minimizes the cloud infrastructure cost.
- It provides low cost cloud services by reducing the infrastructure cost for the cloud vendors.
- It gives the monetary benefit with the large under-utilized infrastructure to the established enterprises.

Table I shows the comparison between data backup and recovery techniques along with the advantages and drawbacks.

Table I Comparison between data backup and recovery techniques

Approach	Advantages	Disadvantages
SBA [13]	• SBA algorithm recovers the data file without any data loss. • Low Central Processing Time (CPU) utilization • Low cost • High data privacy	• With the increase in the data size, there is an increase in the processing time.
PCS [18]	• High reliability • High privacy • Low cost	• High complexity
Bloom Filter [20]	• High data integrity and security	• High encoding and decoding time
HS-DRT [22]	• Illegal data recovery by third party interception becomes almost impossible and an extremely safe data backup system can be realized at reasonable cost. • High cipher code strength and data recovery rate.	• High cost • High complexity
ERGOT system [24]	• Obtains exact match retrieval • High privacy	• High time complexity • High implementation complexity
Linux box [27]	• Simple process • Low implementation cost	• High bandwidth consumption • Low privacy • Complete server backup at a time
CBSRS and HBSRS [30]	• Triggered only when the server failure is detected.	• Cost increases with the increase in the data size
SBRR [35]	• Low cost • Works properly even if the router fails	• Inconsistencies between logical and physical configurations may lead to some performance problem • It is unable to include optimization with the cost reduction
Rent out the Rented Resources [39]	• Cost depends on the utilization of cloud infrastructure	• Implementation is complex • Resources must kept under special attention due to rented concept

III. Factors in Data Recovery

A. *Recovery Time Objective (RTO)*

RTO is the maximum tolerable delay from the time when the services are interrupted due to the server downtime to the time when the services are resumed again. RTO is the most important data recovery factor due to every second in downtime of high service cost in the commercial field. It can range from seconds to days according to the service types [42], data recovery protocols [43-49], CSPs [50] and cloud monitoring system [51]. The types of service decide the kind of appropriate data protection for efficient maintenance and lower operating cost of the cloud system. This is called as service differentiation. It divides all services into higher and lower class according to the Quality of Service (QoS). The higher class involves key protection and lower class involves general protection. The cost of different classes is differentiated.

The protocols focus on the significant reduction in the time and cost required for data recovery. These two factors cannot be achieved simultaneously, as in the dual server cluster data storage technology. Various studies evaluated the recovery protocols to find the best protocol for reducing the recovery time and cost. But, the RTO may vary in different time models proposed by different CSPs. This causes little error in the recovery protocols. Hence, there arises a need to establish a standard model for data recovery.

The cloud monitoring system plays a significant part in the data recovery. The recovery time includes time for detecting the service failure and time for restoring the services. During the occurrence of service failure, there is loss of data files for the services. The backup server does not aware about the loss and does not take any solution. This extends the server downtime. The data recovery is completed quickly as fast as the failure is detected. Heartbeat detection technology is the common method applied for making the main server to send a data message to other server as a periodical backup to inform that the server still works normally. The backup server proceeds to the data recovery operation, if the heartbeat detection message is not received from the main server. The cloud monitoring system should be aware of the occurrence of risk before it causes the server downtime. It is difficult to predict the natural disasters, the warning system monitors the regular environmental changes of the servers, human errors and running state of the machines. The warning system will take solutions for the data recovery issues and notify the backup server to be ready to carry on the operations of the main server. This decreases the failure rate of the main server.

B. *Recovery Point Objective (RPO)*

RPO is the maximum acceptable loss in the data storage during the server failure. The server downtime occurs suddenly without any indications. However, the data backup is done for a fixed time. When the server failure occurs at the time when the new backup is just finished there will be no data loss, as the lost data can be restored from the backup. But, when the failure occurs before the commencement of data backup period, the entire data in the period will be lost. This is the worst scenario. A metastorage [47] is proposed for dividing the big files into smaller ones, as smaller the size of stored data, less time and cost are required for data transfer. It provides accurate data recovery with an appropriate backup time.

IV. Issues in Data Recovery

The main issues associated with data recovery are described as follows

Data privacy: Different clients access the cloud using different login credentials. They are freely permitted to upload the private data on the cloud environment. There arises a need to maintain the data privacy. The data owner should only be able to access the private data.

Server relocation: During data recovery, the server is relocated to the cloud environment for transferring the data from the main server to another server. But, this new location is unknown to the client. The clients obtain the data in the similar way without any indication about the server relocation, such that the location transparency of the relocated server to the clients and third party, while data is transferred to the remote server.

Data security: The client data is stored at a central data repository with comprehensive security. This security protocol should be followed in the remote repository. In the remote data repository, the data should be completely protected such that the unauthorized access of the remote cloud either intentionally or unintentionally by third party or any other client cannot be possible.

Reliability: The remote cloud should own the reliability characteristics. In the cloud computing environment, the main cloud stores the complete data. Each client is dependent on the main cloud for data access. Thus, the data backup plays a reliable role in the cloud. The server should be able to provide data to the client instantaneously either from the main cloud or remote server.

Cost effectiveness: The implementation, backup and recovery cost of the remote server play a significant role for creating the main cloud and correspondent remote cloud. The cost for establishing the remote setup and implementing should be low.

V. Failures in Cloud Storages

The failure in the cloud storage can occur at any time. On June 2010, due to software vulnerability in a Cisco switch, the data centre of Hosting.com had a downtime for a few hours. On April 2011, Amazon reported that Elastic Block

Store (EBS) volumes were out of working. As a result, Amazon's EC2 and RDS services went down for four days. The failures are caused by the accident damage to the applications, natural disaster, etc. Table II shows the classification of these failures.

Table II Classification of cloud storage failures

Reasons	Descriptions	Solutions
Human error	Sudden human faults in the daily operations on the storage systems cause accidental server downtime.	• Human works with intelligent error detection system to complement their own operations, thereby reducing the error rate. • Strengthen manual efforts on the aspect of operation skills training.
Natural characteristics of the components in storage systems and software vulnerability	Every storage system has a Mean Time To Failure (MTTF), so the system may have an unexpected downtime.	The storage system need maintenance regularly and another back-up system will be running instead of the main system in maintenance.
Bad operational conditions	Each storage system needs an appropriate environment conditions including humidity, temperature, altitude, dust, etc. When any of these environment factors is out of tolerable range, there is a rise in the rate of system failures.	A monitoring system for running environment factors is required. The environment of main system need to be adjusted regularly and it need other assistance system to maintain appropriate running environment, such as air conditions, dust remover and so on.
Natural disasters	The location of the storage systems suffers from natural disaster such as earthquake, Tsunami, etc. The system will be breakdown instantly. The disasters are beyond the system handle range for human.	A parallel site that is located far away from the main site in distance is needed. When the main site is ruined, the backup site will be running to compensate for the downtime losses.

VI. CONCLUSION

In this paper, a detailed review of data recovery approaches in the cloud computing domain along with the advantages and disadvantages is presented. These data recovery strategies provide best performances under various circumstances within a short time span. SBA algorithm recovers the data file without any data loss. With the increase in the data size, there is an increase in the processing time. Bloom filter yields high data security. But, it requires high encoding and decoding cost. PCS is comparatively reliable, maintain privacy of each resource and also incurs minimum infrastructure cost. However, it is unable to control the implementation complexities. HSDRT is an efficient technique for the mobile client devices. It fails to manage the low cost for the implementation of the recovery and also not able to control the data duplication. ERGOT system is completely based on the semantic analysis and unable to focus on the time and implementation complexity. Linux Box model uses very simple concept of data back-up and recovery at a very low cost. However, data protection level in this model is very low. SBBR focuses on the implementation and reduction cost. But, it fails to concentrate on the optimization concept and redundancy. With a new virtualization concept, REN cloud also focuses on the low cost infrastructure with the complex implementation and low security level. The Cold and Hot back-up strategy performs backup and recovery during failure detection. However, the cost increases gradually with the increase in the data size. To mitigate the existing issues, our future work focuses on the data recovery through Forward Error Correction (FEC) implemented on the cloud computing environment.

REFERENCES

[1] Z. Ke, W. Hua, and L. Chunhua, "Cloud storage technology and its applications," ed, 2012.

[2] Z. Jian-Hua and Z. Nan, "Cloud computing-based data storage and disaster recovery," in *Future Computer Science and Education (ICFCSE), 2011 International Conference on*, 2011, pp. 629-632.

[3] C. Modi, D. Patel, B. Borisaniya, A. Patel, and M. Rajarajan, "A survey on security issues and solutions at different layers of Cloud computing," *The journal of supercomputing,* vol. 63, pp. 561-592, 2013.

[4] S. Subashini and V. Kavitha, "A survey on security issues in service delivery models of cloud computing," *Journal of network and computer applications,* vol. 34, pp. 1-11, 2011.

[5] J. Wang, Y. Zhao, S. Jiang, and J. Le, "Providing privacy preserving in cloud computing," in *Human System Interactions (HSI), 2010 3rd Conference on*, 2010, pp. 472-475.

[6] M. Mowbray and S. Pearson, "A client-based privacy manager for cloud computing," in *Proceedings of the fourth international ICST conference on COMmunication system softWAre and middlewaRE*, 2009, p. 5.

[7] D. Lin and A. Squicciarini, "Data protection models for service provisioning in the cloud," in *Proceedings of the 15th ACM symposium on Access control models and technologies*, 2010, pp. 183-192.

[8] M. R. Abbasy and B. Shanmugam, "Enabling data hiding for resource sharing in cloud computing environments based on DNA sequences," in *IEEE World Congress on Services (SERVICES)*, 2011, pp. 385-390.

[9] S. J. Stolfo, M. B. Salem, and A. D. Keromytis, "Fog computing: Mitigating insider data theft attacks in the cloud,"

in *Security and Privacy Workshops (SPW), 2012 IEEE Symposium on*, 2012, pp. 125-128.

[10] R. P. Sarang and R. K. Bunkar, "Study of Services and Privacy Usage in Cloud Computing," *International Journal of Scientific Research in Computer Science and Engineering,* vol. 1, pp. 7-12, 2013.

[11] Vishal Kadam and M. Kumbhkar, "Security in Cloud Environment," *International Journal of Scientific Research in Computer Science and Engineering* vol. 2, pp. 6-10, 2014.

[12] A. Bala and Y. Osais, "Modelling and simulation of DDOS Attack using SimEvents," *International Journal of Scientific Research in Network Security and Communication,* vol. 1, pp. 5-14, 2013.

[13] K. Sharma and K. R. Singh, "Seed block algorithm: a remote smart data back-up technique for cloud computing," in *International Conference on Communication Systems and Network Technologies (CSNT)*, 2013, pp. 376-380.

[14] R. Gandhi and M. Seshaiah, "Data back-up and recovery techniques for cloud server using seed block algorithm," *International Journal of Engineering Research and Applications,* vol. 5, pp. 91-95, 2015.

[15] M. Shaikh, A. Achary, S. Menon, and N. Konar, "Improving cloud data storage using data partitioning and data recovery using seed block algorithm," *International Journal of Latest Technology in Engineering, Management & Applied Science,* vol. 4, 2015.

[16] K. Pophale, P. Patil, R. Shelake, and S. Sapkal, "Seed Block Algorithm: Remote Smart Data-Backup Technique for Cloud Computing," *International Journal of Advanced Research in Computer and Communication Engineering,* vol. 4, 2015.

[17] M. Tidke, V. Jadhav, S. Parab, S. Patil, Y. Patil, U. Scholar, *et al.*, "Seed Block Algorithm: A New Approach for Data Back-up and Recovery in Cloud Computing," *International Journal of Engineering Science,* vol. 4093, 2016.

[18] C.-w. Song, S. Park, D.-w. Kim, and S. Kang, "Parity cloud service: a privacy-protected personal data recovery service," in *Trust, Security and Privacy in Computing and Communications (TrustCom), 2011 IEEE 10th International Conference on*, 2011, pp. 812-817.

[19] H. Jung, Y. Park, C.-W. Song, and S. Kang, "PCS: a parity-based personal data recovery service in cloud," *Cluster Computing,* vol. 20, pp. 2655-2668, 2017.

[20] S. Zhang, H. Zhou, Y. Yang, and Z. Wu, "A joint Bloom Filter and cross-encoding for data verification and recovery in cloud," in *Computers and Communications (ISCC), 2017 IEEE Symposium on*, 2017, pp. 614-619.

[21] A. Singh, S. Garg, S. Batra, N. Kumar, and J. J. Rodrigues, "Bloom filter based optimization scheme for massive data handling in IoT environment," *Future Generation Computer Systems,* 2017.

[22] Y. Ueno, N. Miyaho, S. Suzuki, and K. Ichihara, "Performance evaluation of a disaster recovery system and practical network system applications," in *Systems and Networks Communications (ICSNC), 2010 Fifth International Conference on*, 2010, pp. 195-200.

[23] S. Suguna and A. Suhasini, "Overview of data backup and disaster recovery in cloud," in *International Conference on Information Communication and Embedded Systems (ICICES)*, 2014, pp. 1-7.

[24] G. Pirro, P. Trunfio, D. Talia, P. Missier, and C. Goble, "Ergot: A semantic-based system for service discovery in distributed infrastructures," in *Proceedings of the 2010 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing*, 2010, pp. 263-272.

[25] B. Solanki and J. Jha, "Web Service Discovery using Relational Database and Apache Lucene," 2015.

[26] J. Jayalakshmi and G. Mathuramgbigai, "A SURVEY ON BACKUP RECOVERY ISSUES IN CLOUD COMPUTING," 2017.

[27] V. Javaraiah, "Backup for cloud and disaster recovery for consumers and SMBs," in *Advanced Networks and Telecommunication Systems (ANTS), 2011 IEEE 5th International Conference on*, 2011, pp. 1-3.

[28] K. Bangale, K. Nadhe, N. Gupta, S. S. Parihar, and G. Mankar, "Smart Remote Health Care Data Collection Server," *International Journal of Computer Science and Mobile Computing,* vol. 3, pp. 415-422, 2014.

[29] M. Assefi, M. Wittie, and A. Knight, "Impact of network performance on cloud speech recognition," in *Computer Communication and Networks (ICCCN), 2015 24th International Conference on*, 2015, pp. 1-6.

[30] L. Sun, J. An, Y. Yang, and M. Zeng, "Recovery strategies for service composition in dynamic network," in *Cloud and Service Computing (CSC), 2011 International Conference on*, 2011, pp. 60-64.

[31] D. Niu, L. Rui, C. Zhong, and X. Qiu, "A composition and recovery strategy for mobile social network service in disaster," *The Computer Journal,* vol. 58, pp. 700-708, 2015.

[32] M. G. Narke, M. A. Harijan, M. A. Shinde, and H. Sonawane, "A smart data backup technique for cloud computing using seed block algorithm strategy," 2015.

[33] M. Raje and D. Mukhopadhyay, "Algorithm for Back-Up and Recovery of Data Stored on Cloud along with Authentication of the User," in *Information Technology (ICIT), 2015 International Conference on*, 2015, pp. 175-180.

[34] D. Niu, L. Rui, H. Huang, and X. Qiu, "A service recovery method based on trust evaluation in mobile social network," *Multimedia Tools and Applications,* vol. 76, pp. 3255-3277, 2017.

[35] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. H. Katz, A. Konwinski, *et al.*, "Above the clouds: A berkeley view of cloud computing," Technical Report UCB/EECS-2009-28, EECS Department, University of California, Berkeley2009.

[36] S. Vishwakarma and P. D. Soni, "Cloud Mirroring: A Technique of Data Recovery," *International Journal of Current Engineering and Technology,* vol. 5, 2015.

[37] T. Kulkarni, K. Dhaygude, S. Memane, and O. Nene, "Intelligent Cloud Back-Up System," *International Journal of Emerging Engineering Research and Technology,* vol. 2, pp. 82-89, 2014.

[38] S. Agalya, S. Bhavithra, and S. S. Benitta, "AN INTELLIGENT DATA BACK-UP AND RETRIEVING TECHNIQUE FOR CLUSTER ENVIRONMENT," *Journal of Engineering And Technology Research,* vol. 3, pp. 1-9, 2015.

[39] B. Cully, G. Lefebvre, D. Meyer, M. Feeley, N. Hutchinson, and A. Warfield, "Remus: High availability via asynchronous virtual machine replication," in *Proceedings of the 5th*

USENIX Symposium on Networked Systems Design and Implementation, 2008, pp. 161-174.

[40] K. Sharma and K. R. Singh, "Online data back-up and disaster recovery techniques in cloud computing: A review," *International Journal of Engineering and Innovative Technology (IJEIT)*, vol. 2, pp. 249-254, 2012.

[41] K. Keahey, M. Tsugawa, A. Matsunaga, and J. Fortes, "Sky computing," *IEEE Internet Computing*, vol. 13, pp. 43-51, 2009.

[42] M. Wiboonrat, "An empirical IT contingency planning model for disaster recovery strategy selection," in *Engineering Management Conference, 2008. IEMC Europe 2008. IEEE International*, 2008, pp. 1-5.

[43] J. Che, Y. Duan, T. Zhang, and J. Fan, "Study on the security models and strategies of cloud computing," *Procedia Engineering*, vol. 23, pp. 586-593, 2011.

[44] M. Wiboonrat, "System reliability of fault tolerant data center," in *The Fifth International Conference on Communication Theory, Reliability, and Quality of Service, Chamonix, France*, 2012, pp. 19-25.

[45] R. Singha, "A multi-site disaster recovery solution based on ip storage networking," in *International Conference on Information and Computer Networks*, 2012, pp. 139-142.

[46] M. Wiboonrat and K. Kosavisutte, "Optimization strategy for disaster recovery," in *Management of Innovation and Technology, 2008. ICMIT 2008. 4th IEEE International Conference on*, 2008, pp. 675-680.

[47] D. Bermbach, M. Klems, S. Tai, and M. Menzel, "Metastorage: A federated cloud storage system to manage consistency-latency tradeoffs," in *IEEE International Conference on Cloud Computing (CLOUD)*, 2011, pp. 452-459.

[48] O. H. Alhazmi and Y. K. Malaiya, "Evaluating disaster recovery plans using the cloud," in *Reliability and Maintainability Symposium (RAMS), 2013 Proceedings-Annual*, 2013, pp. 1-6.

[49] F. Xiang, C. Liu, and B. Fang, "Novel “rich cloud” based data disaster recovery strategy," *J. Commun*, vol. 6, pp. 92-101, 2013.

[50] T. Wood, E. Cecchet, K. K. Ramakrishnan, P. J. Shenoy, J. E. van der Merwe, and A. Venkataramani, "Disaster Recovery as a Cloud Service: Economic Benefits & Deployment Challenges," *HotCloud*, vol. 10, pp. 8-15, 2010.

[51] G. Aceto, A. Botta, W. De Donato, and A. Pescapè, "Cloud monitoring: A survey," *Computer Networks*, vol. 57, pp. 2093-2115, 2013.

Authors Profile

I. Benjamin Franklin is Assistant Professor in Department of Computer Applications, St. Joseph’s College of Arts & Science (Autonomous), Cuddalore, Tamil Nadu, India. He has 12 years of teaching experience. He is currently pursuing Ph.D. He has published and presented 7 papers in various national/international journals and conferences. His area of interest includes Networking, Cloud Computing and Grid Computing.

T. N. Ravi is Assistant Professor and Research Co-ordinator of PG and Research Department of Computer Science, Periyar E.V.R. College (Autonomous), Tiruchirappalli, Tamil Nadu, India. He has 27 years of teaching experience and 15 years of research experience. His area of interest includes Parallel Computing, Data Mining, Networking and Image Processing. He has guided 30 scholars for M.Phil. and Ph.D. He has published more than 37 research papers in reputed international/national journals.