HOSTED BY
ELSEVIER
CrossMark
Available online at www.sciencedirect.com
ScienceDirect
Karbala International Journal of Modern Science 1 (2015) 122−128
http://www.journals.elsevier.com/karbala-international-journal-of-modern-science/

# Enriched multi objective optimization model based cloud disaster recovery

S. Suguna*, A. Suhasini

*Department of Computer Science and Engineering, Annamalai University, India*

Received 7 August 2015; revised 2 September 2015; accepted 4 September 2015
Available online 28 November 2015

**Abstract**

In cloud computing massive data storage is one of the great challenging tasks in term of reliable storage of sensitive data and quality of storage service. Among various cloud safety issues, the data disaster recovery is the most significant issue which is required to be considered. Thus, in this paper, analysis of massive data storage process in the cloud environment is performed and the massive data storage cost is based on the data storage price, communication cost and data migration cost. The data storage reliability involves of data transmission, hardware dependability and reliability. Reliable massive storage is proposed by using Enriched Multi Objective Optimization Model (EMOOM). The main objective of this proposed optimization model is using Enriched Genetic Algorithm (EGA) for efficient Disaster Recovery in a cloud environment. Finally, the experimental results show that the proposed EMOOM model is effective and positive and reliable.


*Keywords:* Cloud computing; Disaster recovery; Enriched multi objective optimization model (EMOOM); Enriched genetic algorithm (EGA)

## 1. Introduction

The development of an information based society increases the digitized based resources. Because it increases the data storage capacity and storage cost. Additionally, various applications are needed accordingly various storage capacities are also needed in the cloud [1]. Still, the storage space allocated to the cloud applications is often not completely used. The service provider is confronting the exchange between the fast development of information resource and cost control. The cloud service providers not only create enormous essential data by themselves, but also need massive information resources. Instead, huge amount of manpower and equipment storage are required to store the resource information [2]. So, the new data storage device is needed by the cloud service provider that should take features such as dynamically extensible storage capacity, storage virtualization, and reliable data storage. But still, in the cloud storage system have major issues, for example Disaster Recovery (DR).

DR represented (serves) as the desirable feature for entire cloud enterprises. Due to impose of stark tradeoff, the DR adoption becomes limited. The application to the crash point can be recovered by substantial application overhead, minimal geographical

* Corresponding author.
*E-mail address:* sugunaresearch14@gmail.com (S. Suguna).
Peer review under responsibility of University of Kerbala.

http://dx.doi.org/10.1016/j.kijoms.2015.09.001

separation, financial considerations, among the recovery and primary sites [3]. To overcome these cloud disasters, there are various authors using the different kinds of recovery techniques such as seed block algorithm [4], HS-DRT [5], Party Cloud Services [6], ERGOT [7], Linux Box [8] to recover the data. There are numerous companies which are generated the DR techniques as needed for their business stability. Data is stored in cloud infrastructure so as to recover the data loss when the disaster occurs. The other main problem is security and quality during the data recovery in the cloud. There are several security issues such as access control [9], authentication [10], confidentiality [11] and so on.

Thus, in this paper, analysis of massive data storage process in the cloud environment is performed and the massive data storage cost is based on the data storage price, communication cost and data migration cost. The data storage reliability involves the data transmission, hardware dependability and reliability. Reliable massive storage is proposed by using Enriched Multi Objective Optimization Model (EMOOM). The main objective of this proposed optimization model is using Enriched Genetic Algorithm (EGA) [12] for efficient Disaster Recovery in a cloud environment. Finally, the experimental results show that the proposed EMOOM model is effective and positive and reliable.

## 2. Related works

In Ref. [13] the author considered the problems of outsourced data in the cloud data storage like non storage and retrieving the data from servers and observed the lacking in SPs (Service Provider) during the outsourcing. This author proposed a technique of distributing the data among multiple cloud storage models for providing the equal data distribution among the market SPs. The distributed data storage provides better privacy and also reliability among the customers in the market. In this data fragmentation technique is used for the distribution of data.

The cryptography technique provides decryption and encryption of the user data authentication information to protect it from the unauthorized attacker or user. This approach makes sure that the high cloud storage integrity, improved easy identification and error localization of misbehaving cloud sever is to improve the cloud storage performance by using remote data integrity checking concept. Thus, in Ref. [14] author proposes data storage method in a cloud environment with less computation cost and time by using cloud data centers. The cloud severs permit the users to store their data on the cloud without taking care of integrity and correctness of the data.

Cloud computing is a model for authorizing proper on demand network access model to a shared pool of structural computing resources that can be quickly released and provisioned with the least management service or service provider [15]. By means of capacity to provide users a shared resource, dynamically scalable over the internet [16]. Cloud computing has recently improved as a promising hosting platform that implements a usage of a collection of infrastructure, application, service consisted pools of network, computer, storage resources, and information. Therefore, along with these benefits storing a huge amount of data as well as critical information on the cloud.

Data backup and disaster recovery is an important problem with cloud computing. These problems arise at the time of cloud service provider or network failure. The data backup offered by the many cloud service providers is very costly and more expensive for the SMB's and customers. Thus, in Ref. [17] author proposes a method such as online data backup for the cloud in term of disaster recovery. This method minimizes the cost of the solution and also protects the cloud data from disaster. Additionally, it provides migration process for one service provider to another cloud service provider. This proposed solution calculates the dependency of the cloud consumers of service providers. Additionally estimate the data backup cost.

In Ref. [18] author proposes a Knowledge as a Service (KaaS) framework for disaster cloud data management (Disaster-CDM). The main objective of this proposed work is to store huge volumes of data from various sources, simplifying the searching operation and supporting their integration and interoperability. A combination of NoSQOL and Relational database are used to store the cloud data. In Ref. [19] authors present a Disaser Tolerance (DT) method that covers all layers of Infrastructure as a Service (IaaS). The technical aspect of this proposed work is the enhancements of Qemu-kvm hypervisor and Seven-stage DT algorithm, Eucalyptus cloud, Libvirt virtualization toolkit, and the Amazon Elastic Compute Cloud (EC2) public API, Eucalyptus cloud and tools [20] proposed that load balancing which was done based on the genetic algorithm. During the load balancing, map reduce algorithm assign the rank to each input file. Then the input files are optimized by applying the genetic algorithm these will lead to reducing the number of node size and increases the efficiency of the system.

## 3. Enriched Multi Objective Optimization Model (EMOOM)

The Enriched Genetic Algorithm (EGA) [21] is one of the optimization techniques. In this proposed work, the EGA is used to optimize the cloud storage environment to avoid the disaster in the cloud. The EGA operation can briefly explained as Initialization, Evaluation, Coding, Crossover, Mutation, Reproduction and finally terminating condition [22]. The proposed Enriched Multi Objective Optimization Model (EMOOM) model provides enough space for cloud storage optimization to accomplish adequate advantages. The EMOOM model can effectively select the resources from different kind of cloud providers to store backup data. Thus, the DR cost as low as possible and also DR time remains as minimize as possible when disaster occurs. However, this process, ensuring the data reliability. The proposed work designs a multi-cloud based DR Service model perspective of the Service Provider (SP). Initially, collect and gather complete cloud files, Then the EMOOM allow a SP to influence the resources from other cloud SP that create a data DR service to fulfill the user's needs, increase its revenue and to improve its market reputation. To guarantee high data reliability, the EMOOM adopts EGA with 3 replica data redundancy process. When directing data backup process, the DR SP (service provider) will logically select the resources from multiple cloud SPs including itself and this process is relied upon a particular optimized scheduling strategy. The storage cost, data recovery speed, network communication cost is used to make a decision for optimized storage. Such that it can accomplish two important optimization goals such as shortening data recovery time and reducing data backup cost as much as possible.

In general, the EMOOM model comprised of data DR Service users and multiple cloud SPs. In Fig. 1, the $CP_1$ (disaster of service provider) signifies the data DR service provider. All the users of $CP_1$ which can be individuals or other cloud SPs or even other enterprises. They have the proper privileges and account of $CP_1$. The $CP_2$ and $CP_M$ are represented as other cloud SPs that provide cloud resources to $CP_1$. Each $CP$ cloud interface is the cloud storage interface which sends or receives data to or from users. Data backup request of $CP_1$ is given and queued in the request buffer of $CP_1$ that delivers the request in one time period which are waiting for optimized cloud storage scheduled.

Initially, the Replica Scheduler (RS) gets requests from Request Buffer (RB), it creates three replicas and each replicas reports them to those CPs (cloud provider). The Resource Manager is responsible for observing the changes of resource utilization of all CPs. The Metadata is a cloud database; it comprises the information about CPs' resource usages and replicas' locations.

## 4. Optimized data scheduling by Enriched Genetic Algorithm

In the proposed architecture, the optimized scheduling strategy of $CP_1$'s RS is the most significant, which regulates the effect of EMOOM model. The responsibility of optimized scheduling approach in EMOOM is to minimize or shorten the data recovery time and data backup cost as much as possible.

The cloud data scheduling using the Enriched Genetic Algorithm has made the process more powerful and efficient. The data are considered as solutions and those solutions are taken as an initial population. The success of a genetic algorithm is based on the selection of appropriate fitness value [23] based on which the data (chromosomes) are evaluated. Based on the fitness value of the data, the retaining of data is decided. The improper calculation of fitness value leads to decreasing the processing time of genetic algorithm. Then higher fitness value data are taken to the next level of assessment. Next is selection operator through which best fit data are identified. Each data has its own probability which can be found by the sum of all data probability. The data which are having best fit value attains more preference. Then the crossover operator combines two or more data which produce effective new data. This process is necessary in genetic algorithm because it supports the process to select the data to the next level. Mutation produces the changes in the old set of chromosomes (data) which introduces a simple change in each data. This process is used to find the eligible new data in the search space. Once the highest fitness value data are found, and then the data are arranged based on the size of it. EGA initial population generation process is shown in the below Table 1.

Enriched Genetic Algorithm (EGA) flowchart is shown in Fig. 2.

Fitness value is calculated by using the following equation (1)

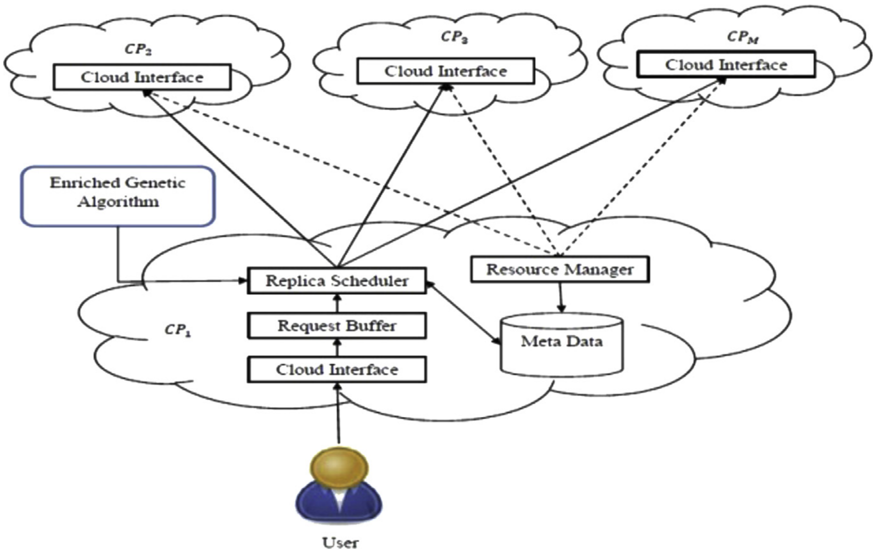


Fig. 1. EMOOM architecture.

$$F = \min\left\{max\{ck\} + \sum f(d_i)\right\} \quad (1)$$

where $ck$ represents the kthe node chromosome
$d_i$ represents that distance between chromosome

In cloud, data are stored continuously on a Virtual Machine (VM) location. First tasks are scheduled and related resources are allocated to the data which is stored in the cloud. So that the data from the cloud can be accessed at any time.

## 5. Data recovery process

During the data recovery procedure, the following steps describe the execution flow.

Table 1
EGA initial population.

| EGA initial population |
|---|
| Step 1: Initialization Set POP = popsize |
| Step 2: Execution Set Chromosome Number = 0 |
| Step 3: If Chromosome Number = popsize |
| Stop the process |
| Show the initial population Else |
| ChromosomeNumber = ChromosomeNumber+1and set POP[ChromosomeNumber] |
| Step 4: Set geneNumber = 0 |
| Step 5: If geneNumber = n goto Step3 |
| Step 6: set geneNumber = geneNumber+1 |
| Step 7: Execute the process and select the values such that n > geneNumber |

1. A data recovery request is the initial step from the user to the data disaster recovery service provider $CP_1$. This request is transmitted via cloud interface.
2. The second step is an immediate authentication process once the request reaches the $CP_1$. It authenticates the user's details like privilege and account; and neglect the request once it is found to be not the legal one.
3. In order to find the backup locations of the data, every $CP_1$ has its own recovery manager that found the data in the Meta database.
4. If the $CP_1$ find the data in the database in three different locations, then it compares all the location and selects the one with highest recovery bandwidth.
5. Consider the selected location be $CP_{1x}$, and the recovery manager of $CP_x$ alerts the $CP_{1x}$ to made the temporary access authorization which can be used by the recovery proxy of the user who send the request in begin to read the replica. Once the authorization started, the $CP_x$ informs $CP_1$.
6. The recovery proxy of the user who gave request receives the authorization information and data location from the $CP_1$ and the data from the $CP_x$ is taken from its recovery proxy and intimates $CP_1$.
7. Once the data is retrieved from the $CP_{1x}$, it abolish the temporary access authorization which was crested before according to the instruction of $CP_1$. Then it records the information regarding the user request for later use.

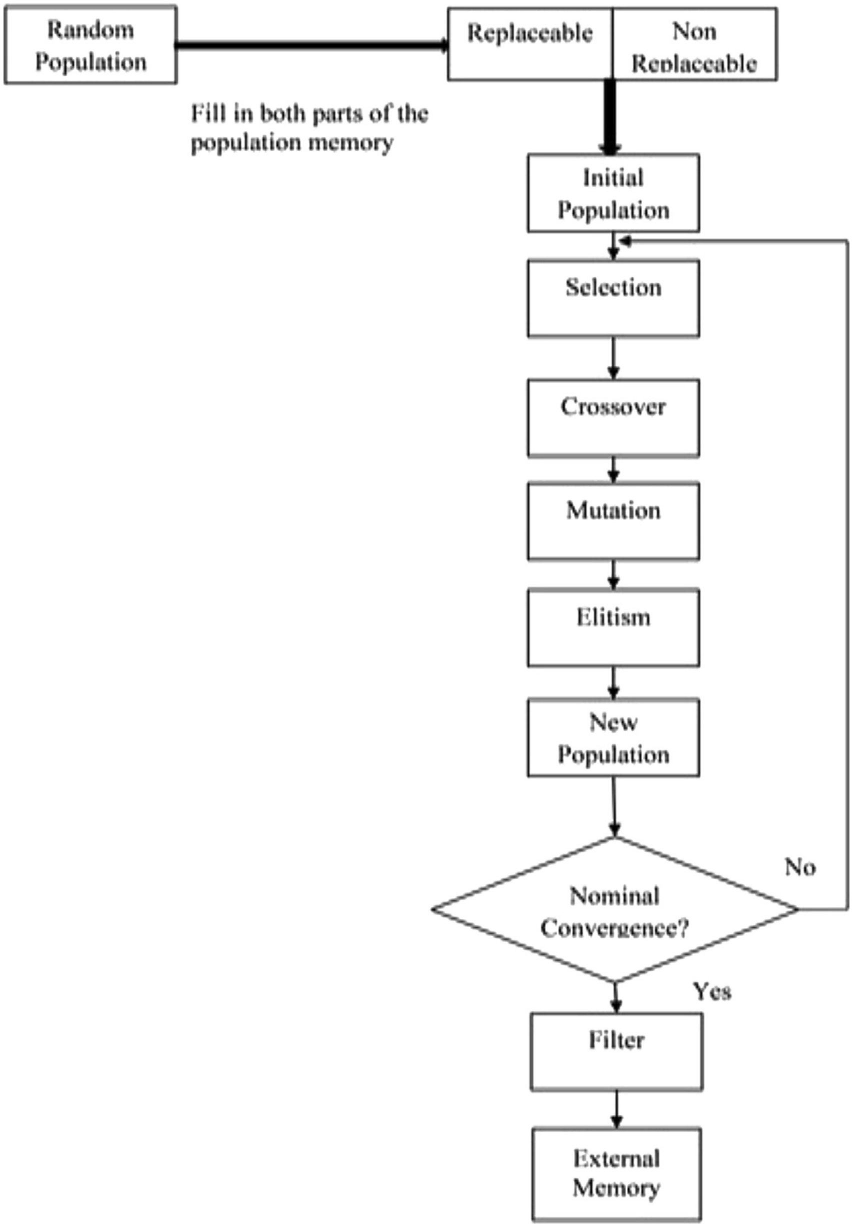


Fig. 2. Enriched genetic algorithm flowchart.

8. In parallel, the $CP_x$ analyze and monitor the traffic consumption of the request to charge the $CP_1$ when needed.

## 6. Experiment design and results analysis

In this proposed work, the main remote server and cloud's server storage is kept 5 GB and 8 GB respectively, which can be extended as per the requirements. The remote server kept more memory space to compare with the cloud's server storage because the additional data information is stored on remote server only. Experimental parameters are shown in Table 2.

One of the main advantages of EGA is the random guided search, which is needed to test the search space by using search point (fitness value). This proposed

Table 2
Parameter.

| Parameter | Cloud server and remote server | |
|---|---|---|
| CPU | I3 Processor | |
| Memory | Cloud server 5 GB | Remote server 8 GB |
| OS | Windows, Linux | |
| HDD | 500 GB | |
| Number of data centers | 20 | |
| Number of VM | 500 | |
| CPU Speed/VM | 1, 3, 3.5 and 4 GHZ | |
| Number of tasks | 15, 20, 25, 30, 60, 80, 100 | |

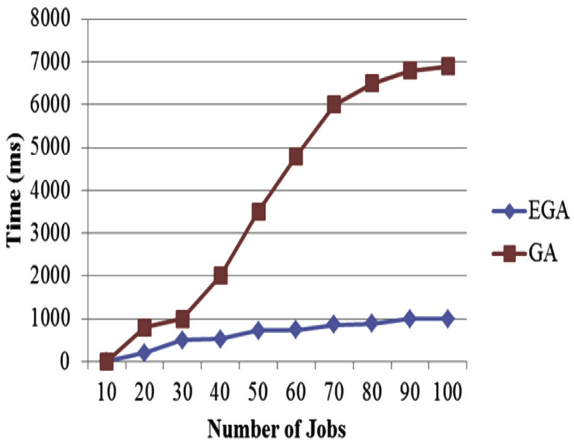


Fig. 3. EGA and GA comparison.

work looks for the local nearest optimum value, which maps the nearest data value allocated on the cloud.

Fig. 3 shows the comparison between proposed Enriched Genetic Algorithm and exiting Genetic Algorithm (GA) in term of scheduling Time (ms). The proposed EGA shows the promising results during comparison.

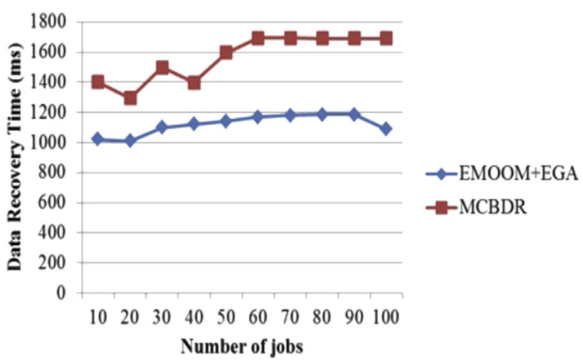


Fig. 4. Data Recovery time.

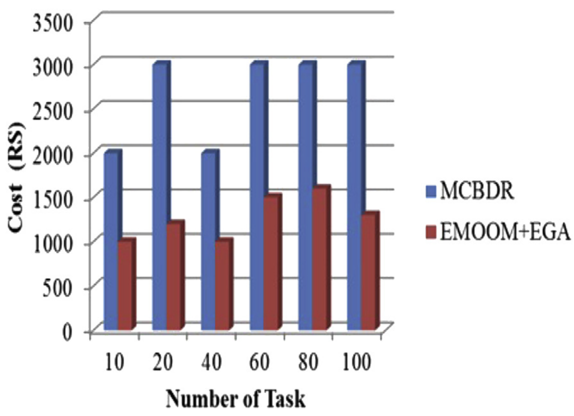


Fig. 5. Overall data backup cost.

Fig. 4 shows a comparison between the proposed EMOOM model with EGA and Existing Multi-Cloud Based Disaster Recovery Service (MCBDR) model in term of data recovery time. The proposed model takes minimum time for data recovery process by making use of EGA. The data backup cost consists of network communication cost, data storage cost, and data migration cost. Therefore the overall data backup cost for scheduling period is denoted by COST and formulated as

$$COST = \sum_{j=1}^{N} \sum_{i=1}^{M} \left( S_j{}^*D_j{}^*R_{ij}{}^*SP_i + S_j{}^*E_{ij}{}^*TP_i \right) \qquad (2)$$

where, $S_j$ represents the store duration at j th task data size.
$R_{ij}$ represents the data replicas count
$E_{ij}$ is denoted as Boolean variable
$SP_i$ service provider

Where $R_{ij}$ represents the data replicas count stored in $CP_1$ for take $T_j$, whereas $E_{ij}$ is denoted as Boolean variable. The Boolean variable is set as 1 when $CP_i$ comprise the task $T_j$ data replica otherwise it should be 0 (Fig. 5).

## 7. Conclusion

This paper presents a reliable massive storage by using Enriched Multi Objective Optimization Model (EMOOM). The main objective of this proposed optimization model is using Enriched Genetic Algorithm (EGA) for efficient Disaster Recovery in a cloud environment. Multiple cloud SPs can be used cooperatively by the data DR service provider. The cloud users only required to deal with the SP, utilizing very unified and simple cloud service interface, without taking care of the internal processes among heterogeneous cloud. EMOOM can ensure low backup cost, short recovery time and high data reliability by utilizing intelligent Enriched Genetic Algorithm for data scheduling. This proposed method is suitable for different sorts of data DR process. The results given by this proposed system are better than exiting approach used previously.

## References

[1] H. Timothy Wood, Andrés Lagar-Cavilla, K.K. Ramakrishnan, PipeCloud: using causality to overcome speed-of-light delays in cloud-based disaster recovery, SOCC, ACM, 2011, pp. 27—28.

[2] Feng Liu, Haitao Wu, Xiaochun Lu, Xiyang Liu, Lei Fan, Genetic algorithm based optimization model for reliable data storage in cloud environment, Adv. Sci. Technol. Lett. 50 (2014) 74—79.

[3] O.H. Alhazmi, Y.K. Malaiya, Assessing disaster recovery alternatives: on-site, colocation or cloud, in: IEEE 23rd International Symposium on Software Reliability Engineering Workshops (ISSREW), 2012, pp. 19—20.

[4] R.V. Gandhi, M. Seshaia, A. Srinivas, C. ReddiNeelima. Data back-up and recovery techniques for cloud server using seed block algorithm", Int. J. Eng. Res. Appl., 5, 2.

[5] Kruti Sharma, Kavita R. Singh, Online data back-up and disaster recovery techniques in cloud computing: a review, Int. J. Eng. Innovative Technol. (IJEIT) 2 (5) (November 2012).

[6] Phalke, Onkar Kulkarni, Abhijit Fulsagar, Harish Rane, Ajinkya Chothave, Literature survey on cloud data recovery techniques, Int. J. Adv. Found. Res. Comput. (IJAFRC) 2 (January 2015). Special Issue (NCRTIT 2015).

[7] Tanay Kulkarni, Krupali Dhaygude, Sumit Memane, Onkar Nene, Intelligent cloud back-up system, Int. J. Emerg. Eng. Res. Technol. 2 (7) (October 2014) 82—89.

[8] Kolipaka Kiran, Janapati Venkata Krishna, Smart data back-up technique for cloud computing using secure erasure coding, Int. J. Comput. Trends Technol. (IJCTT) 16 (3) (Oct 2014).

[9] Ramandeep Kaur, Pushpendra Kumar Pateriya, A study on security requirements in different cloud frameworks, Int. J. Soft Comput. Eng. (IJSCE). ISSN: 2231-2307 3 (1) (March 2013).

[10] R. Bala chandar, M.S. Kavitha, K. Seenivasan, An approach for broader range of security in cloud computing, Int. J. Secur. Priv. Trust Manag. ( IJSPTM) 2 (1) (February 2013).

[11] S. Suganya, S. Dhanalakshmi, Evaluation of disaster recovery in cloud computing, Int. J. Multidiscip. Res. Dev. 2 (6) (June 2015) 300—304.

[12] A. Kaleeswaran, V. Ramasamy, P. Vivekanandan, Dynamic scheduling of data using genetic algorithm in cloud computing, Int. J. Adv. Eng. Technol. 5 (2) (2012) 327.

[13] B. AmarNadh Reddy, P. Raja Sekhar Reddy, Effective data distribution techniques for multi-cloud storage in cloud computing, Int. J. Eng. Res. Appl. 2 (5) (2012).

[14] Jahnvi S. Kapadia, Survey on various techniques for data storage security in cloud computing, Int. J. Sci. Eng. Res. 4 (5) (2013).

[15] Qian Wang, Cong Wang, Kui Ren, Wenjing Lou, Jin Li, Enabling public auditability and data dynamics for storage security in cloud computing, in: European Symposium on Research in Computer Security (ESORICS), IEEE, 2009.

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

[17] K. Grolinger, M.A.M. Capretz, E. Mezghani, E. Exposito, Knowledge as a service framework for disaster data management, in: IEEE 22nd International Workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE), 2013, pp. 313—318.

[18] M.C. Caraman, S.A. Moraru, S. Dan, C. Grama, Continuous disaster tolerance in the IaaS clouds, in: 13th International Conference on Optimization of Electrical and Electronic Equipment (OPTIM), IEEE, 2012, pp. 1226—1232.

[19] R. Dharani, M. Narmatha, Secured data sharing with traceability in cloud environment, Int. J. Invent. Comput. Sci. Eng. 1 (8) (2014).

[20] R. Palson Kennedy, T.V. Gopal, An effective optimized genetic algorithm for scalable information retrieval from cloud using big data, J. Comput. Sci. 10 (Issue 6) (2014).

[21] Z.Q. Zhang, J.X. Zhou, N. Zhou, X.M. Wang, L. Zhang, Shape optimization using reproducing kernel particle method and an enriched genetic algorithm, Comput. Methods Appl. Mech. Eng. 194 (39—41) (1 October 2005).

[22] Susmi Routray, An enhanced genetic algorithm approach to ATM network design, Int. J. Inf. Technol 3 (1) (2010).

[23] S. Sanjay, S. Pradeep, V. Manikanta, S.S. Kumara, P. Harsha, Genetic algorithm based approach for the selection of projects in public R&D Institution, Indian J. Comput. Sci. Eng. (IJCSE) 2 (Sep 2011).