Genetic Algorithm for Load Balancing in Cloud Computing

Shweta Dwivedi[1], Dr Saket Agrawal[2]
[1]*Research Scholar, Rajshree Institute of Management and Technology, Bareilly*
[2]*Dean, Rajshree Institute of Management and Technology, Bareilly*

ABSTRACT - A structural design in which virtual machines are implicated and connect to the cloud service provider is called cloud computing. On the behalf of the users, the virtual machines connect to the cloud service provider. The uncertainties overload the virtual machines. The genetic algorithm is implemented for the migration of virtual machine in the earlier study. The genetic algorithm is low depicts latency within the network is high at the time of virtual machine migration. The genetic algorithm is implemented for virtual machine migration in this study. The proposed algorithm is applied in MATLAB in this work. The obtained results are compared with the results of earlier algorithm. Various parameters like latency, bandwidth consumption, and space utilization are used to analyze the achieved results.

Keywords- Cloud Computing, Genetic Algorithm, Virtual Machine, Migration

I. INTRODUCTION

The cloud computing can be described as an on-demand service pool which connects various servers to each other for providing services to aiming clients. The cloud providers may contain direct access to these services. Therefore the resources can be used according to the requirement. The user can extracts and modifies the data stored in the clouds. The different services to the user are provided on demand using a feature called "cloud service provider". This trait makes certain that the amount of services being utilized for any number of times can be employed for calculating the expense of the user to access that service. The cloud computing system provides extremely complicated applications in different environments. In addition, some skilled concentrated services are provided in each environment. In cloud computing, common group of resources is provided to the users. Using cloud computing, the users can utilize these resources according to their need everywhere. The main objective of this technology is to maintain the minimum cost to access the services. It is analyzed that the software and hardware assets obtained using internet remain present in the virtual system and supports to provide the services. The user accesses a common group of resources using cloud computing on the basis of demand. The virtualization allows user to subscribe and use the services for a certain time period by getting access of the common group of resources using cloud computing. The cloud computing reduces application cost reduced and allows the user to access more hardware parts [1]. The information regarding physical place or the design of the system is not required in this technology. Some of the significant aspects of cloud computing are identified as the geographic allocation based service orientation, homogeneity, virtualization, current security and inexpensive software. In cloud computing, the user can use applications without installing. The user can also manage and get the access of the private files at any location just using internet. In cloud computing, the centralized storage aspect provides competent bandwidth, memory, computing and resource distribution. The load balancing is a technique using which the whole load of the network is divided and allocated to the several resources or devices. This approach improves the competence of resource employment and enhances the response time of any task as well. Moreover, this approach avoids the condition in which some centers are under loaded while others are overloaded. The trustworthiness and accessibility of the data can be improved with the help of redundancy by using various components for load balancing. The load is measured within the system in terms of the network load, memory usage, and CPU load [12].

1.1 Types of load balancing

For classifying the load balancing, following two methods are applied:

A. Static: The current information of the system is not needed by this technique. This technique works on the earlier information. In order to establish this cloud service provider, the homogeneous resources are utilized in this method. In addition, the overall resources do not show flexibility within the static environment. The memory, performance, capability and processing power are needed on the basis of the need. The changes are rejected in such situations in runtime. This technology is not very competent in the heterogeneous environments. This technology can be adopted easily in the static environments. The resources are distributed in this environment on the basis of first-come-first-server strategy. They may be classified as:

- **Sender initiated**: In this situation, the sender starts the procedure of load balancing.

➔ **Receiver initiated:** In this situation, the receiver starts the load balancing procedure.

B. Dynamic: This approach uses the current state of the system and does not need earlier information. In this environment, Individual resources are established in the clouds. The resources are dynamic in nature in this environment. This approach does not need previous information and just uses the execution time information.

➔ **Symmetric**: In this case, both sender and receiver begin the load balancing process.

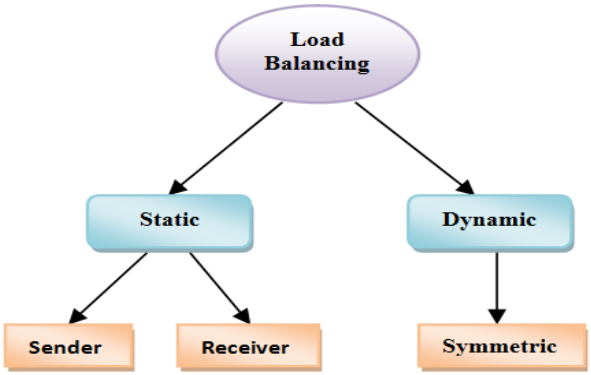

Figure 1.4: Types of load balancing

1.2 Existing Load Balancing Algorithms

In the past, a lot of load balancing algorithms have been designed [13]. A brief description of some of these algorithms is given below:

- **Round Robin**: This algorithm allots time to every process in a periodic way likewise the context switching of processor. Therefore, the same processor time is provided to all the processors in this algorithm. In cloud computing, round-robin method is applied to distribute same quantity of load across various nodes. Equal amount of load may not be available in all the nodes with same competence due to the variation in the capability of the nodes. This algorithm becomes more effectual in case of equally capable nodes. This technique is more frequently utilized for HTTP request processing due to the similarity of requests.
- **Weighted Round Robin:** This algorithm is the advanced version of round robin algorithm. This algorithm removes the issues being faced in the earlier algorithms. Some weight is allotted to each node. The suitable numbers of requests are received on the basis of these assigned loads. The weight is allotted to each instance according to the potential of the processing. This aspect is extremely important as it determines the way in which the instance behaves. A load is distributed to each server existing in the environment. This gives information about the inactivity of a server and distributes the load of two to the noticed server. The IP balancer is connected to the above two dominant servers using two requests. The other two less performing servers are allotted with all the requests. Each node receives some traffic during the allocation of similar weights to the total nodes.
- **Central Queuing:** Central queuing algorithm is retained in the effective load balancing at the central manager. This phenomenon handles all the activities queuing on VM. The manager performs action for processing if it is identified in the ready state. Else ways, it is ignored till identifying the suitable resource for it's in the virtual machine.
- **Min-Min Algorithm:** This algorithm is a popular scheduling algorithm. During load balancing, this algorithm is applied successfully. This algorithm allots execution time to all the unfinished operations. These operations are ordered on the basis of the execution time. In this algorithm, the operation having minimum execution time is selected first. After selection, the processing is carried out. Moreover, the execution time is computed again. The whole process is repeated till the completion of all tasks.
- **Honey Bee Foraging Algorithm:** This algorithm is developed on the basis of the behavior of the honey bees. This algorithm performs tasks in the same way as the honey bees and high-quality sources as VM. The precedence of bees if not denoted here as tasks can be determined using the beehive. All the VMs are sorted and precedence is implicated to the ability tasks on the basis of resources. The task is allotted to the maximum ability VM. This algorithm communicates in two directions. After completing task, the information is sent back to the beehive. Then, the next priority task is allotted. The overall procedure continues till the completion of all tasks [14].
- **Genetic Algorithm:** The genetic algorithm is applied in the dynamic environment and the soft computing technique. The GA algorithm shows better performance as compared to the First Come First Serve or Round Robin algorithms. The GA algorithm provides an enhanced search space that is implemented to the intricate objective function. This algorithm avoids local optimum answer by getting trapped.

The three imperative steps of GA are described below:

- **Selection:** The preliminary population is selected in the random manner.
- **Crossover:** The fitness pair of individuals is identified for the overlapping.
- **Mutation:** This step depicts the mutation value as the small probability value and toggling of bits from 0s to 1s and 1s to 0s. The novel group of individuals that are prepared for overlapping are applied as output.

II. LITERATURE REVIEW

Rajwinder Kaur et al. (2012) stated that the information was exchanged and several advantages were provided to the clients by the cloud computing. Based on the utilization of provided resources, the charges were paid by the clients [21]. The Cloud computing system stored information and circulated resources in open environment. The measurement of stored information was done in a speedy way in the open environment. Therefore, the load balancing was an important topic in cloud computing. The dynamic workload over different centers was provided by the load balancing. This method also ensured the non- overloading of individual center. Further, this approach made sure the authorized usage of resources. This method also enhanced the execution of the system. Different types of promising loads like memory, CPU and system load occurred in distributed computing. Load balancing was known as a procedure of detecting overloaded centers and shifted further load to various centers.

Soumya Ray et.al (2012) stated that load balancing in cloud computing had become a main concern due to the enormous expansion of technology. Various modern algorithms provided enhanced task scheduling for resource allocation [22]. For providing benefits in terms of extra income with sophisticated load balancing algorithms, the resources should be used effectively. In this work, various load balancing methods were highlighted. The main purpose of this work was to find out the computational tools to perform simulation in cloud computing system. These tool were later adapted as input. This phenomenon was the main point of execution inspection of load balancing algorithms.

HU Baofang et.al (2012) introduced a new task scheduling algorithm in cloud computing system. A modified adaptive genetic algorithm was proposed in this work [23]. The proposed approach was completely based on the priority method. The proposed task scheduling guaranteed the minimal execution time and the QoS requirement of client's behavior. In this study, an integrated powerful feature was designed on priority basis for representing the modified object. The proposed approach provided benefits of condensed iterative task and smaller iteration time duration. A comparison among the performances of proposed approach and various other scheduling algorithms was carried out. The proposed approach showed better concurrence rate as per the obtained results.

Punithasurya et.al (2013) stated that the security measures should be taken necessarily while dealing with public cloud. The security was the main factor in the cloud computing [24]. Validation, approval and access control were involved in the security. In the cloud storage, various access control techniques occurred. The main part of cloud computing was identified as access control. This indicated that just the legal clients could have the access of services offered by the cloud computing system. Security was the key concern in cloud computing. For security purpose, a method named Role Based Access Control (RBAC) was employed. This technique could improve the time, site and accessibility.

Vimmi Pandey, et.al (2013) reviewed the mobile token in dynamic application. This application was used in mobile phones and produced a code with the help of OTP (One Time Password). The produced OTP code could be used only one time for login [25]. A method of OTP was explained in this work. There were two step included in this method. These steps are identified as Registration step and Login step. At first, the client registered itself by filling details in the registration form. The client entered the Login screen in the next step. The login display produced OTP for login purpose. The OTP code was achieved using three parameters e.g. recent time, 4-digit PIN code and Init-code. This produced code could be used only for a particular period of time (generally 2 or 3 minutes). This aspect ensured security against the snooping attacks and man-in-middle attack. Thus, OTP code was identified in a safe way.

Ankur Mishra et.al (2013) presented a review of two approaches for providing security to cloud computing environment. These approaches were recognized as Virtualization and Multi-tenancy. In cloud computing environment, these approaches provided security [26]. The third coalition preserved information. This factor provided Saas and PaaS which were vital for the security. The security operations besides those lines were performed using Virtualization and Multi-tenancy schemes. The major purpose of Virtualization was to create physical pc capability in spite of the fact that amid more than two computers, every computer was non-physical or virtualized. There were two classes in which visualization was divided. These classes were called Para virtualization and Full virtualization. The two designs of virtualization were called Hosted and Hypervisor. In addition, Multi-tenancy was defined as the capability to provide access of structure service to numerous users

frequently as base and code base. Multi-tenancy can be connected the single levels. These levels were known as application level, middleware level, working structure and device level. Further, security of virtualization and multi-tendency approaches had been reviewed in this work.

Wei Guo et.al (2013) stated that the unlimited storage potential, computing capacity and data services was provided by the cloud computing [27]. The cloud computing came out as a new application field for both an individual and business enterprises because of these features. The capability of data center was limited. Therefore, several questions regarding the mode of data slice placement in suitable data hub were rose. This became an important part and affected the capability of the field. The association charges between traced data segments were measured cautiously by the data placement system in this work. This event reduced the distributed operation costs to the most likely level, mainly the price changes among different distributed operations. Similarly, in addition, this technique raised universal load balancing issue curing in the data centre. This technique travels a long distance related to the genetic algorithm and ensured that the technique could offer efficient data placement solutions in a speedy way. This technique understood the universal load balance problems in a better way as per the obtained results. This approach decreased the charge of circulated association up to 10% than other existing approaches.

Ruhi Gupta, et.al (2014) recognized cloud computing as an immensely growing technology of the present and the future [28]. Various techniques were analyzed in this work for eliminating the problem of load balancing from the cloud computing system. The growing techniques for load balancing and their effects on the cloud computing system were also analyzed in this work. In future, the distributed approach would be improved further for load balancing to reduce the workload of the virtual machine.

Tingting Wang, et.al (2014) identified task scheduling as a main severe issue of the cloud computing system. A huge amount of data and clients were present in the cloud computing. The allocation of resources distribution requests and their reuse became an important aspect [29]. Therefore, effective task scheduling techniques were required to satisfy the demands of the clients and to guarantee the improved utilization of resources. This event smartened the general overall performance of the cloud computing network. The new attributes of cloud and genuine adaptive genetic algorithm (AGA) was presented to eliminate this problem. This algorithm was one of the advanced scheduling algorithms. This algorithm was mostly dependent on the adaptive algorithm of double-fitness, task across time and load balancing genetic algorithm (JLGA). This method provided services for task scheduling series with lesser and common movement strategy. Also, this method provided inter-nodes load balancing. Greedy algorithm was applied in this work to initialize the population. This algorithm considered difference in populace to explain the load intensiveness among nodes and weights multi-fitness function. Further, the AGA algorithm with JLGA algorithms was compared for performance analysis using simulations. The validity of scheduling algorithm was proved along with the efficacy of the advanced technique as per the simulation outcomes.

III. RESEARCH METHODOLOGY

For handling the problems related to node failure in cloud networks, an algorithm named BFO is used in this research. Several nodes are included in a proposed algorithm. Depending upon the failure rate and least execution time, a candidate node is chosen among all these nodes. In this scenario, the threshold value is fixed using the master node. There are two parameters included in this threshold value. These are the maximum execution time and failure rate. The nodes which have equal to or less failure rate and least execution time are chosen as the candidate nodes by the master node. In comparison to threshold value, the value of node N1 is less. Thus, this node is chosen as a candidate node. There is a less and a higher parameter included in node N2. Therefore, it is not possible to select this node as candidate node. A node N3 is elected as candidate node since it includes the value equivalent to threshold. Another node N4 similarly cannot be chosen as candidate node since its value is higher than that of the threshold value. After it is selected, the candidate node starts its functioning. In this scenario, various tasks are initiated. One node moves from its location once the task is completed. Thus, task failure occurs as a result. For remove the issue of failure arising because of node mobility, a new methodology is proposed. In the proposed algorithm, a new parameter known as master node time is included. The ultimate time connecting the end users is known as master node time which helps in performing node cooperation. Following are the formulas given for calculating the master node time:

1. E-cost= maximum execution time + Time taken by the master node (master node time)

After that we will calculate profit of each node.

2. Profit of each node = E-cost+ Failure node of each node

3. Weight of each node= No. of tasks + maximum execution time/ Profit

The selection of node which has maximum weight is done. The given formula measures the weight.

Following are the steps to be performed by proposed algorithm:

Step 1: Get list of all VMs working on all hosts.
Step 2: Initialize no migration is done.
Step 3: Get resource utilization, failure rate, and execution time of all machines.
Step 4: Built transition matrix for hosts and VMs.
Step 5: Loop will execute until all machines on over utilized hosts are migrated.
Step 5.1: Calculate the current utilization of each host for that particular VM that needs migration.
Step 5.2: Check creation history of the VM.
Step 5.3: Compare increase in utilization of selected hosts with other hosts.
Step 5.4: Select host for which increase in utilization is minimum End loop
Step 5.6: If maximum utilization exceeds upper utilization threshold go to step 5.1.
Step 6: Else choose that particular host for migration.
Step 7: return migration List
END

IV. RESULT AND DISCUSSION

MATLAB simulator is used to implement the proposed algorithm since in real time scenarios, its complexity is high. Based on power usage and execution time, the comparison among the performance of proposed and existing algorithms is evaluated. Table 1 shows the simulation parameters used in this research.

Table 1: Simulation Parameters

Number of VM	10
Number of cloudlets	60
Host Memory	2 GB
Processor	Xenon
Number of Data centers	5

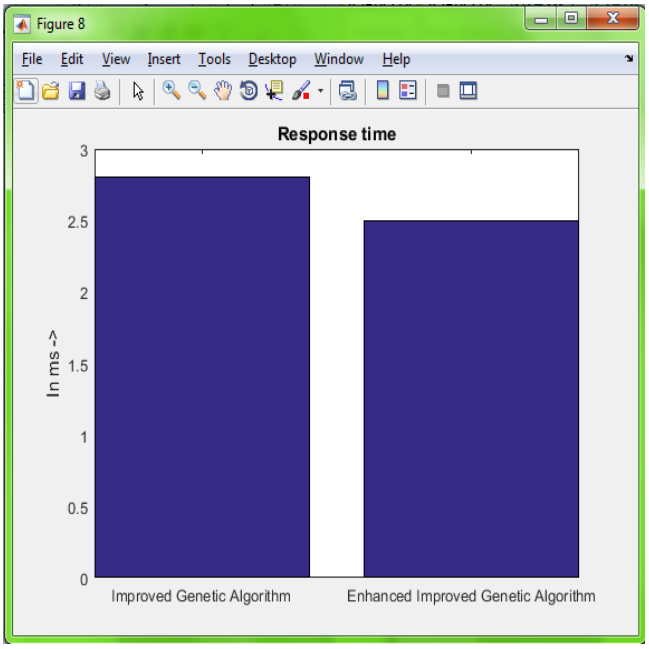

Fig 1: Comparison graph of Response Time

Figure 1 shows the response time of the improved genetic algorithm and proposed enhanced improved genetic algorithm compared for the performance analysis. The response time of enhanced improved genetic algorithm is less as compared to improved genetic algorithm.

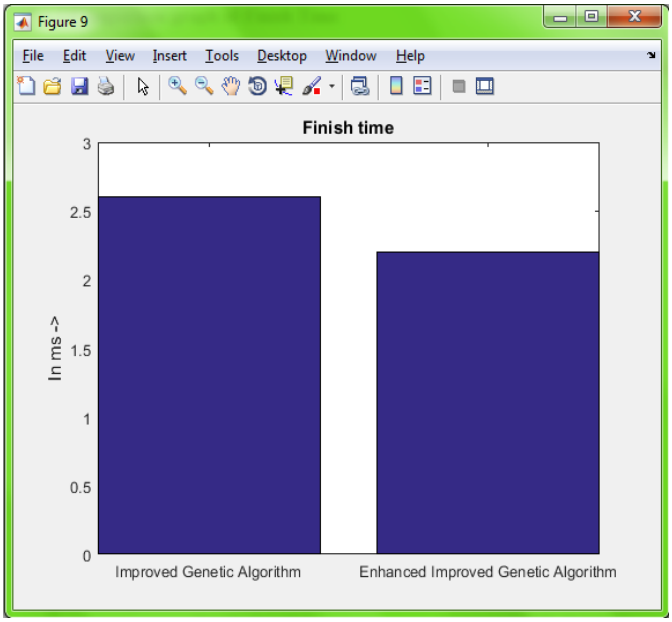

Fig 2: Comparison graph of Finish Time

Figure 2 shows the finish time of the improved genetic and proposed enhanced improved genetic algorithm compared for the performance analysis. The finish time of the enhanced improved genetic algorithm is less as compared to improved genetic algorithm.

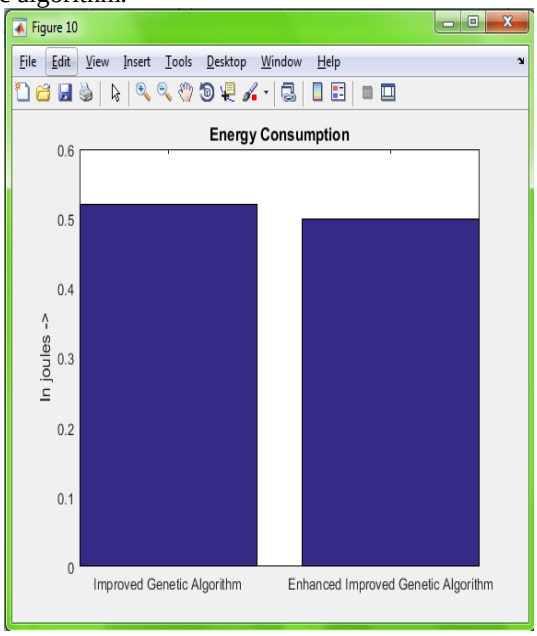

Fig 3: Comparison graph of Energy Consumption

Figure 3 shows the energy consumption of the improved genetic algorithm and proposed enhanced improved genetic algorithm compared for the performance analysis. The energy consumption of enhanced improved genetic algorithm is less algorithm has less cost as compared to improved genetic algorithm.

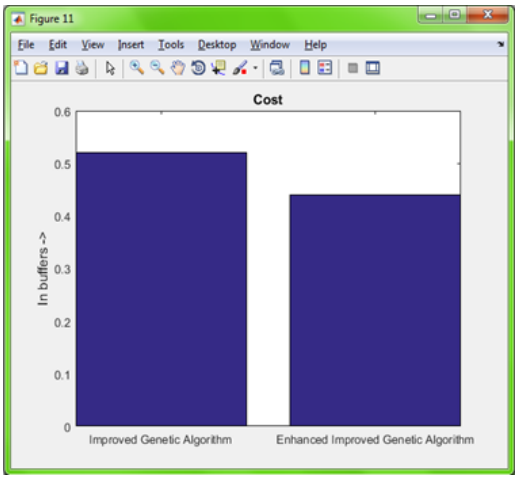

Fig 4: Comparison graph of Cost

Figure 4 shows the cost of the improved genetic algorithm and enhanced proposed improved genetic algorithm compared for the performance analysis. The enhanced improved genetic

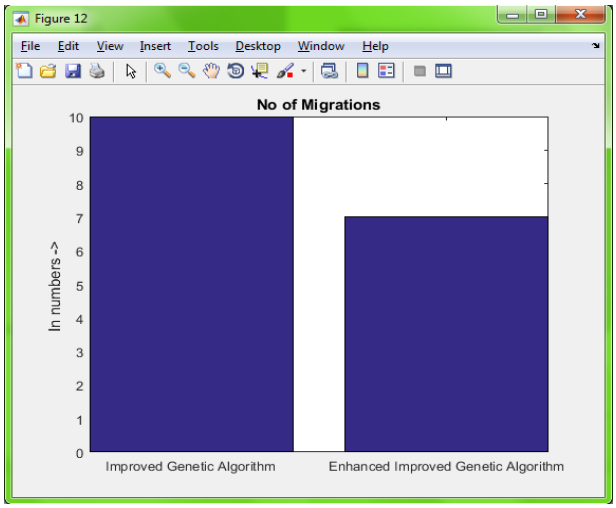

Fig 5: Comparison graph of No of Migrations

Figure 5 shows the number of migrations of improved genetic and proposed enhanced improved genetic algorithm compared for the performance analysis. The number of migration of enhanced improved genetic algorithm is less as compared to improved genetic algorithm.

V. CONCLUSION

This work focuses on load balancing issue being faced in the cloud architectures. In the systems, latency can be increased due to load balancing. For performing virtual machine migration, the previous research has applied genetic algorithm. It is seen through this research that the complexity of genetic algorithm is high. Thus, the time of virtual machine migration is increased. This research work aims to perform virtual machine migration by applying improved Genetic technique. MATLAB is used to implement the proposed algorithm and several parametric values are calculated to analyze the performance of this algorithm. The results show that in comparison to existing algorithm, the performance of proposed algorithm is better. The future work of this research can be extended by proposing a new security algorithm through which the virtual channel attack of clouds can be isolated.

VI. REFERENCES

[1] Randles, M.; Lamb, D.; Taleb-Bendiab, A., "A Comparative Study into Distributed Load Balancing Algorithms for Cloud Computing", IEEE 24th International Conference on Advanced Information Networking and Applications Workshops (WAINA), April 2010, pp.551-556.

[2] Fang, Yiqiu, Fei Wang, and Junwei Ge, "A task scheduling algorithm based on load balancing in cloud computing", Web Information Systems and Mining. Springer Berlin Heidelberg, 23 October 2010, Pages 271-277.

[3] Chaczko, Zenon, Venkatesh Mahadevan, Shahrzad Aslanzadeh and Christopher Mcdermid, "Availability and load balancing in cloud computing", International Conference on Computer and Software Modeling, Singapore, Vol. 14, 2011.

[4] Samrat Kumar Dey, Md. Raihan Uddin, Kh. Mohaimenul Kabir and Md. Mahbubur Rahman, "Enhancing the Security of Cloud Computing: Genetic Algorithm and QR Code Approach", Proceedings of the 2017 4th International Conference on Advances in Electrical Engineering, 2017.

[5] Shui Han, Jianchuan Xing, "Ensuring Data Storage Through A Novel Third Party Auditor Scheme in Cloud Computing", IEEE computer science & Technology, 2011 pp 264-268.

[6] Sean Carlin, Kevin Curran, "Cloud Computing Security", International Journal of Ambient Computing and Intelligence, January 2011, pp 14-19.

[7] Peter Mell, Timothy Grance, "The NIST Definition of Cloud Computing", Recommendations of the National Institute of Standards and Technology, September 2011, Special Publication 800-145.

[8] Dr Nashaat el-Khameesy,Hossam Abdel Rahman, "A Proposed Model for Enhancing Data Storage Security in Cloud Computing Systems", 2012, vol-3.

[9] Arora Pankaj, Wadhawan C.Rubal, Er.Ahuja P.Satinder, "Cloud Computing Security Issue in Infrastructure as a Service", International Journal of Advance Research in Computer Science and Software Engineering, 2012.

[10] Habib, S. M., Hauke, S., & Ries, S., "Trust as a facilitator in cloud computing: a survey", Journal of Cloud Computing , 2012, page 01-18.

[11] WANG Bei, LI Jun, "Load Balancing Task Scheduling based on Multi-Population Genetic Algorithm in Cloud Computing", Proceedings of the 35th Chinese Control Conference, July 2016, pp 27-29.

[12] Srinivas.J, K. Venkata Subba Reddy, Dr. A. Moiz Qyser, "Cloud Computing Basics", International journal of advanced research in computer and communication engineering , 2012, pp. 343-347.

[13] Tushar Desai, Jignesh Prajapati, "A Survey of Various Load Balancing Techniques and Challenges in Cloud Computing", International Journal of Scientific & Technology Research, November 2013, Volume 2, Issue 11.

[14] Sanjoli Singla, Jasmeet Singh, "Cloud Data Security using Authentication and Encryption Technique", International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) , July 2013 Volume 2, Issue 7, pp 2232-2235.

[15] Bhavna Makhija, VinitKumar Gupta, "Enhanced Data Security in Cloud Computing with Third Party Auditor", International Journal of Advanced Research in Computer Science and Software Engineering, 2013, pp 341-345.

[16] Parsi, K., & M.Laharika, "A Comparative Study of Different Deployment Models in a Cloud", International Journal of Advanced Research in Computer Science and Software Engineering , 2013, pp 512-515.

[17] Barron, C., Yu, H., & Zhan, J., "Cloud Computing Security Case Studies and Research", Proceedings of the World Congress on Engineering, 2013, Volume II.

[18] Gowri G,Amutha.M, "Cloud Computing Applications and their Testing Methogology", International Jouranl of Innovative Research in Computer and Communication Engineering, 2014, pp. 2906-2914.

[19] Wang jibin, zhao Zhigang xu , zhang Hu,Li liang,Hu and Guo ying, "I-sieve: An Inline High Performance Deduplication System Used in Cloud Storage", Tsinghua Science and Technology, 2015, pp. 17-27.

[20] Jibin Wang, Lili Yang, Hu Zhang, Zhaogang Xu, Ying Guo, "PROXY: A High Cost-Performance Cloud Storage Gateway", Advanced Cloud and Big Data 2015 Third International Conference on, pp. 197-204, 2015

[21] Kaur, Rajwinder and Pawan Luthra, "Load Balancing in Cloud Computing", Proceedings of International Conference on Recent Trends in Information, Telecommunication and Computing, ITC, 2012.

[22] Soumya Ray and Ajanta De Sarkar, "Execution Analysis of Load Balancing Algorithm in Cloud computing Environment", International Journal on Cloud Computing: Services and Architecture (IJCCSA), October 2012, Vol.2, No.5.

[23] HU Baofang, SUN Xiuli, LI Ying, SUN Hongfeng, "An Improved Adaptive Genetic Algorithm in Cloud Computing", 13th International Conference on Parallel and Distributed Computing, Applications and Technologies, 2012.

[24] Punithasurya K, Esther Daniel, Dr. N. A. Vasanthi, "A Novel Role Based Cross Domain Access Control Scheme for Cloud Storage", International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), 2013, Volume 2, Issue 3, March 2013, pp 942-946.

[25] Vimmi Pandey, "Securing the Cloud Environment Using OTP", International Journal of Scientific Research in Computer Science and Engineering, 2013, vol-1, Issue-4.

[26] Ankur Mishra, Ruchita Mathur, Shishir Jain, Jitendra Singh Rathore, "Cloud Computing Security", International Journal on Recent and Innovation Trends in Computing and Computation, 2013, pp 36-39.

[27] Wei Guo, Xinjun Wang, "A Data Placement Strategy Based on Genetic Algorithm in Cloud Computing Platform", 10th Web Information System and Application Conference, 2013.