Genetic Algorithm for Load Balancing in Cloud Computing

Arun Gupta[1] & Dr. Anoop Sharma[1]

[1]Department of Computer Applications, Singhania University in Pacheri Bari, Jhunjhunu (Raj.), India
(arun.gupta@gmail.com, sharmaanoop001@gmail.com)

Abstract*—Cloud computing (CC) is a structural layout where virtual computers are involved and connected to the cloud service provider (CSP). The virtual computers establish a connection with CSP on the users' behalf. VMs are overloaded by the uncertainty. The load may be caused by the CPU, memory, or network. In the preceding work, the genetic algorithm was used to migrate virtual machines (VMs). When a virtual machine is moved, the low-latency genetic algorithm shows a high-latency network. In this study, the evolutionary algorithm is used to migrate virtual machines. In this work, the suggested algorithm is utilized in MATLAB. The findings achieved are contrasted with those of an earlier algorithm.*

Acknowledgement. *Words cannot express my gratitude to Dr. Anoop Sharma my guide & my mentor who always stands besides me to support in writing this paper. I am also grateful to my cousin Sanjeev Gupta & my wife Sushma Gupta for their trust in me & for giving me emotional and moral support throughout my journey of this research.*

Keywords: *Cloud Computing, Virtual Machine Migration, Genetic Algorithm, Load Balancing*

1. INTRODUCTION

CC environment provides users the appearance that they have unrestricted computational control. Depending on their individual demands, users can increase or decrease their asset utilization, including their use of energy, but they are unaware of how this control was acquired. When cloud computing is used in computer technology, the consumer is not concerned with the method of calculation and how effectively it operates. It is believed that the cloud itself is an entirely virtualized environment [3]. The infrastructure for data measurement processing and application development can be viewed as a framework unto itself, alongside the database. It offers robust and simple computing facilities; technological advancements like cloud grids and clusters each tend to continue offering access to a significant amount of computing power from a fully virtualized infrastructure by easily consolidating requirements and ensuring a single overarching framework. Additionally, utility computing services are offered through cloud technology. A commercial strategy for providing computing resources in response to consumer requests is described by the term "on-demand computing" [4]. Based on how much service is used, customers pay service providers. All members of the public have access to traditional public services, specifically water, electricity, and gas.

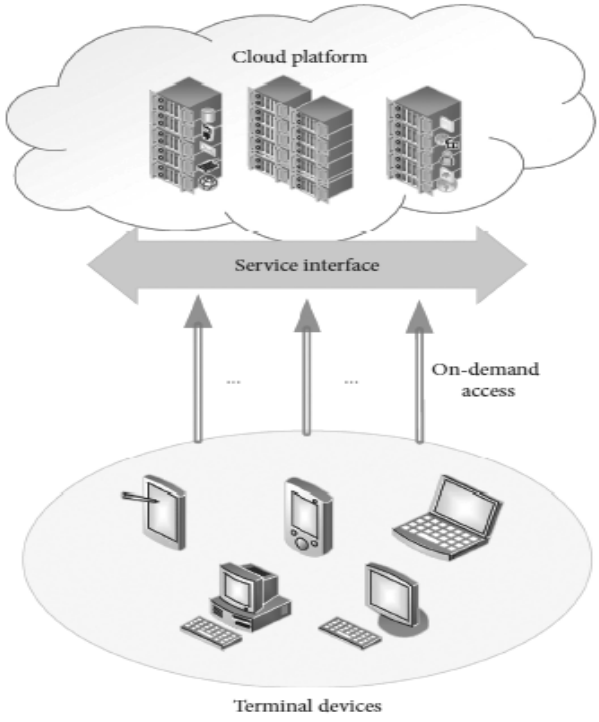

Figure : The structure of cloud computing

Cloud computing's main goal is to support widespread, suitable, and OD network access for computing resources, which can be configured easily, including storage, networks, platforms, and services in shared pools. It is feasible to give and release resources fast and with little managing effort or SP interface [5]. CC framework is depicted in Figure 1. Due to its

pay-as-you-go pricing model, which enables service providers to profit from the cost of renting resources in the cloud without investing in infrastructure, cloud computing does not require a large upfront investment. Resources are managed and leased in a flexible manner. According to the demands of the users, it is possible to release the resources in the cloud. In order to relieve strain on centre load and energy consumption and cut expenses, service providers can proactively release idle resources in the cloud when service demand is low. The data resources obtained by a cloud data centre can be gathered and analysed by the infrastructure providers. Data analysis is used by service providers to identify prospective business trends and control the growth in demand for their services, which is the foundation for them to expand their service footprint and scale. Through devices like mobile phones, PCs, and PDAs, cloud services are easily available over the Internet [6]. A contemporary paradigm for offering computing services that satisfy the changing needs of information technology is cloud computing. Because it can greatly cut expenses and offer support for the management of information technology systems without any issues, it is believed to be of tremendous benefit to consumers and companies. Over the past few years, it has shown to be beneficial in practically every area of industry, including financial trading, e-commerce, academia, and many others. Cloud computing also lowers the absolute risk for capital investments made by IT organizations. Companies that provide cloud services can better manage cloud resources and help with controlling and coordinating revenue and income [7]. The ease of use of computing resources and software, as well as the usage of minimal IT infrastructure, are valued by researchers. The way that data processing services operate can be drastically altered by the cloud. Cloud-based data processing technologies are becoming commonplace in the market area. However, remote data processing is often overseen by a remote provider (CSPs). The list highlights a number of advantages of traditional computer systems, including their unlimited data storage capacity, great performance, and low cost, to name just a few. However, people are worried about losing personal information in contrast. Similar to cloud computing, users are provided with resources in the form of tools and services; they can select the ones that best suit their needs. In cloud environments, users cannot track the origins of their data, where it has been collected from, who has accessed it, or how it has been altered.

1.1. LOAD BALANCING IN CLOUD COMPUTING

A fundamental problem with distributed frameworks and cloud computing paradigms is load balancing. The load balancing process involves dividing workloads among a collection of reliable infrastructure servers. Inconvenient computing involvements, like QoS, and SLA breaches, untrustworthy data processing, and lack of sensitivity, may result from the workloads given without the use of load balancing solutions. It is essential to use load balancing techniques in cloud computing models due to the rise of high-tech computing systems [8]. Infrastructure as a Service is the cornerstone of CC architecture. The cloud data centre makes the implementation of several physical hosts to provide customer services. As the residual resources of each physical host are changing at rapid rate, the impossible process is of implementing the task to the physical host having an enormous amount of resources every time. Instead, let the tasks that users request be arrayed on a physical host that is chosen at random each time [9]. In case the volume of resources requested for the work are found above to the quantity of resources left over on the chosen physical host, the physical host is unable to manage the task. The deployed task will be regarded as failing in this case. The time required to complete a task will be comparably long if the resources requested for the task have similarity with the rest of the resources of the physical host that is chosen for task execution. Because of the unbalanced computational outcomes that occur when a cloud data centre receives successive work requests, users cannot be served in a timely and efficient manner [10]. Due to a variety of factors, the cloud center's data load can become so unbalanced that it is unable to provide users with effective external services. In fact, cloud computing data centres normally guarantee the service performance by allocating jobs in accordance with the respective hosts' maximum load demands. Because of this, the vast majority of physical hosts are largely inactive, wasting CPU resources.

2. LITERATURE REVIEW

B. B. Wang, et al (2020) suggested a method called QuickN to assist in searching the nearest neighbour on the encrypted data on unreliable clouds [11]. Moreover, an optimized algorithm was implemented on the peak of an enhanced search algorithm to save the communicating overheads of a client and no information was leaked during this process. H. Zhu, et

al (2020) discussed that FL (Federated Learning) was utilized for preventing the leakage of secret information [12]. A MOEA (multi-objective evolutionary algorithm) was put forward to optimize the structure of the NN (neural network) of FL with the objective of mitigating the communicating costs and ER (error rates). The DNN (deep neural network) was made more effective on the basis of a scalable method.

Z. L. Jiang, et al (2021) analysed that an ML (machine learning) technique known as FL was utilized for preserving the privacy so that the algorithms were trained on the distributed datasets [13]. A secure NN (neural network) was put forward in (Federated Learning) and the multiple keys were created. Moreover, this approach made the deployment of a DTE (double-trapdoor encryption) technique. The fundamental intend of clients was that EL algorithms were uploaded to the primary CS. Thereafter, the subsequent server was employed to decrypt the initial one. For this, one trapdoor is utilized. T. Kong, et al (2020) recommended a MTD (moving target defence) system called ConfigRand which was able for preventing the information outflows which the shared kernel caused in the CBC [14]. MTD concept was assisted in formulating a framework to generate, distribute and exploit the deceiving system configurations. An innovative technique was put forward for creating these configurations and quantifying their heterogeneity. Z. Wu, et al (2019) investigated a Boolean formula as a set of DFAs and an innovative method adopted for operating an encrypted DFA [15]. The cloud was useful to process these automatons. Three kinds of queries called: conjunctive, disjunctive, and Boolean helped in making the system more efficient, adaptively secure, and mitigating the leakage.

B. H. K. Chen, et al (2018) established a novel CypherDB method that was implemented subsequent to encrypt the complete subcontracted database and execute the queries across this data [16]. The joint data paths were employed to optimize the computing effectiveness. The accessing of database and executing the query led to avoid the information leakage. S. Gao, et al (2020) presented a new CP (ciphertext-policy) known as TrustAccess having reliability and system used to control the attribute hiding access in accordance with blockchain [17]. The system was accessed reliably using this method and its applicability was proved to preserve the safety of policy and attribute. An OHP-CP-ABE (optimized hidden policy ciphertext-policy-ABE) was presented to secure the policy and fulfil the requirements of vast universe access.

S. V. Usov, et al (2019) introduced a TaC-RBAC (Task-Controlled Role-Based Access Control) algorithm based on the restricted guidelines to authorize the limit role and perform the activation legacy. The notion of few privileged user was implemented and the redundancy of user permissions was alleviated [18]. The least number of tasks were executed using the potential of person. The major task was to determine whether multiple roles or a single role activated the least privilege theory or not in a session. The generalized RBAC (Role-Based Access Control) focused on organizing a hierarchy and granting the permissions by higher-level roles. Moreover, TC instructions were considered to perform the permission heritage and the role activation in this algorithm such that it was not possible to allocate the tasks for senior roles to junior roles; it was not essential to allocate the permissions from senior to junior roles; and considering any session at which the allocation was not done, the activation of role was not done.

Y. Wang, et al (2018) presented two rules in which initial emphasized on restricting the association among the task role. The so-called "priority junior roles rule," were executed to determine the capability of a junior position for carrying out a duty, the senior role cannot be allocated [19]. The second rule was utilized to restrict the relation of approval with task for limiting the permissions. Thirdly, only a customer was allowed a function to allocate the job to that role. The appropriateness of the RBAC (Role-Based Access Control) algorithm was proved for organizational systems for handling the information and users via the same entity. In case of controlling the data by a third party, RBAC's access control resolutions were ineffective to fulfil the demands due to the files management after encrypting the third party to client's data privacy.

Likewise, the limitation of RBAC systems was the knowledge of security administrators earlier regarding all potential user agreements and functions. An access provide was utilized to restrict the little versatility as a consumer. R. Ghazal, et al (2020) analysed that the complex task was to change an access privilege of user with no change of its assigned roles [20]. Moreover, this approach focused on maintaining and assisting the users in connecting the roles and other ties among privileges and responsibilities in every solution to control the access. K. Soni, et al (2019) presented a technique

and conduct its computation. The results exhibited that the allocation of those functions was required for enforcing the (Role-Based Access Control) system, and there was no change in accessing the privileges without any change in the roles for protecting the records of unreliable entities [21].

K. Lee, et al (2020) suggested an IBE (Identity-based Encryption) method. This method was utilized to encrypt the data at which the redundant information for recognizing the user was comprised in customer profile [22]. A sender was responsible for encrypting data. For this, the key of the Recipient user was implemented, and the private key generator (PKG) was the trusted party to generate a corresponding personal key to decrypt the data. A public master key was granted by PKGs, and any key was allowed to compute the key for matching that identity with its public master key. The effectiveness of the suggested method was proved due to the known identities of user to an encrypted which was capable of decrypt the data earlier. Likewise, no intender user was required for granting its public key to perform the encryption after extracting the identity from the public key.

Z. -Y. Liu, et al (2021) established a HIBE system which was a modified form of an IBE (Identity-based Encryption) for decentralizing the process to create PK (private keys). This system was planned on the basis of diverse PKGs (private key generators) arranged in a hierarchical structure [23]. The PKG root at the next hierarchical phase led to create only a private key. Moreover, a PKG was not only assisted in generating a secret key for the entities stated. This system allowed series ID to define a user as well as tuple of IDs are executed for comprising the predecessor IDs in the order. To illustrate, a user placed below to the root in the order was described by integrating the strings ID1, ID2, and ID3. Moreover, the established system implemented to publish a cryptographic signature at one level and affect the private keys at the next level.

RESEARCH METHODOLOGY

In this study, an algorithm known as BFO is employed to address issues with node failure in CC. An algorithmic solution is proposed that contains several nodes. A candidate node is selected from amongst all motes based on the failure rate and the smallest possible execution time. In this case, the master node is used to fix the threshold value. This threshold value takes into account two variables. As participant nodes, the master node selects nodes with an execution time that is as short as possible and a comparable failure rate. Unlike the threshold value, node N1's value is lesser. As a result, the selection of such kind of mote is done as a participant mote. Node N2 contains a lower and a greater parameter. Node N3 is chosen as the participant node because it contains a value equal to the threshold. N4 is not selected as a participant node due to its value which is found greater. The candidate node begins to function after being chosen. Many tasks are started in this scenario. Once the task is finished, one node switches to another location. As a result, task failure happens. A novel methodology is suggested to address the problem of failure brought on by node mobility. A fresh parametric quality called master node time is included in the new algorithmic method. Master node time, that aids in collaborating a node, is the actual time that connects the end customers. The formulas provided for determining the master node time are as follows:

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

PSEUDO CODE OF PROPOSED SYSTEM

```
Begin
Input: Virtual machine
Output: Task migration
Define Number of Tasks as Tk
Threshold value of failure rate as FR
Threshold value of execution time as ER
Repeat while virtual machine is selected for the Task (Tk)
 If (FR of machine i> FR of machine i+1)
 If (ER of machine i> ER of machine i+1)
        Select i+1 as best machine
    End if
        End if
End of while
If (virtual i get overloaded=true)
    Calculate weight ()
        If (weigh of i>weight of i+1)
            Select machine i for migration
        Else
```

Select execute weight algorithm

End if

End

3. RESULT AND DISCUSSION

The new approach is implemented in this work using MATLAB as a simulation tool because real-world scenarios would be highly difficult without it. The introduced algorithm is compared with the existing ones with regard to power utilization and execution time. The simulation settings utilised in the presented work are displayed in Table 1.

Table 1: Simulation Parameters

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

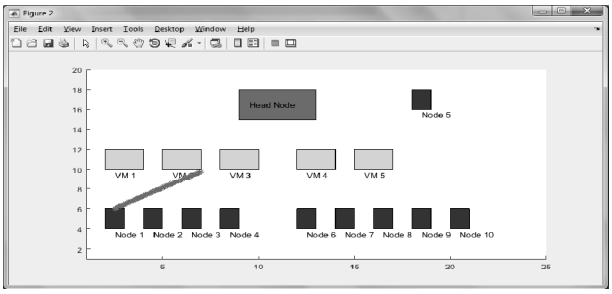

Figure 1: Virtual Machine Migration

Figure 1 illustrates how the best virtual machine is chosen as the one to which the task will be transferred for execution.

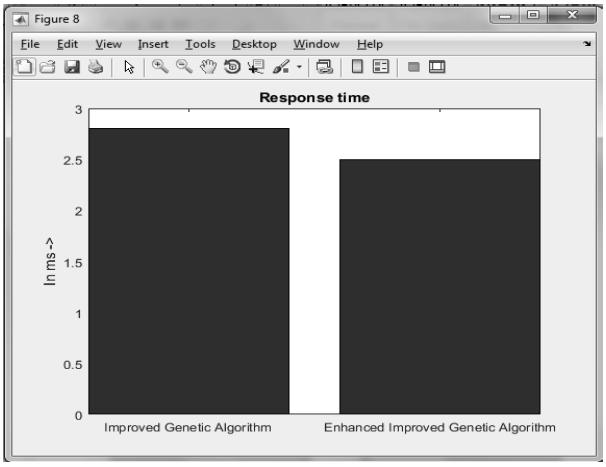

Figure 2 Comparison graph of Response Time

In order to perform comparison of the improved Genetic Algorithm and the new enhanced improved GA, Fig. 2 shows a reaction time-based comparison between both. The new algorithm is better since it responds faster than its predecessor.

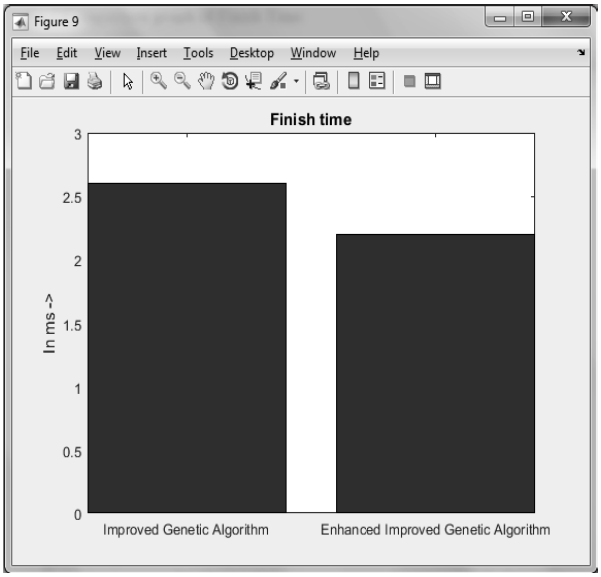

Figure 3 Comparison graph of Finish Time

In order to compare the performances of the existing algorithm with the new enhanced Genetic Algorithm, Fig. 3 shows a finish time-based comparison between the two algorithms. The new algorithm is better because it completes in less time than its predecessor.

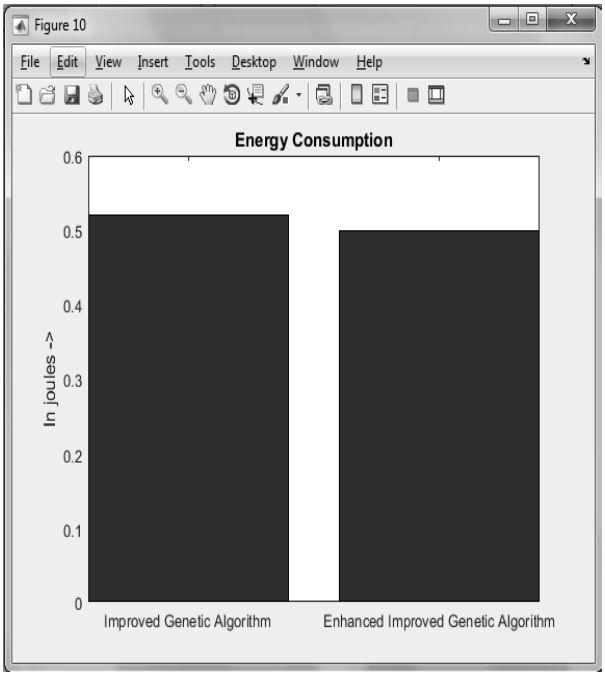

Figure 4: Graph of Energy Consumption Comparison

The Fig. 4 demonstrates the comparison of the introduced GA with the existing algorithm concerning

energy consumption. Because it uses less energy than its predecessor, the new algorithm is superior.

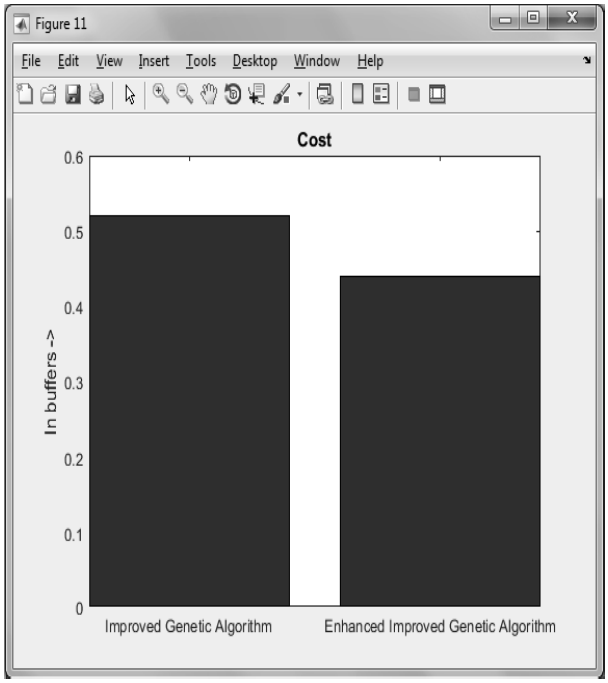

Figure 5 Comparison graph of Cost

To analyse the performance of both of these algorithms, Fig. 5 represents the comparison of the introduced GA with the existing algorithm concerning cost. The new algorithm is better since it costs less than its predecessor.

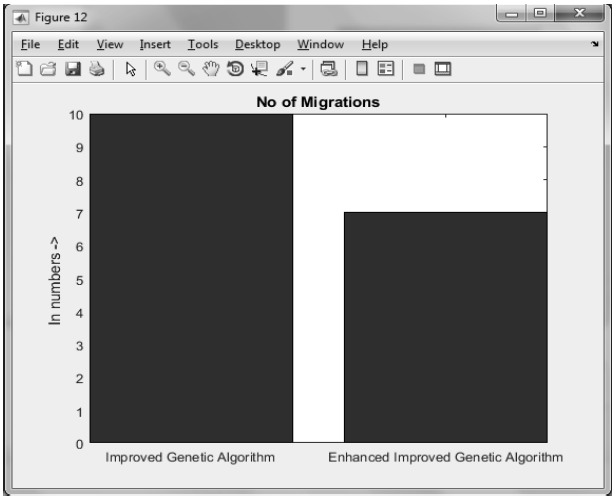

Figure 6: Comparison graph of No of Migrations

On the basis of the number of migrations, Fig. 6 contrasts better GA with new enhanced improved GA. The new algorithm outperforms its predecessor with fewer migrations.

4. CONCLUSION AND FUTURE WORK

The terms "dynamic" and "static" refer to two main groups of load balancing techniques. The choice about load shifting is not made depending on the state of the algorithm's static type scheme at the time. Here, knowledge of the system's resources and applications is necessary. This study aims to find a solution to the load balancing problem that cloud architectures are now experiencing. Due to load balancing, latency in the systems may rise. The work done in the past has applied GA for VM migration. This study demonstrates the high level of genetic algorithm complexity. As a result, virtual machine migration takes longer. The goal of this research project is to migrate virtual machines using an enhanced genetic method. The newer algorithm is implemented in MATLAB, and numerous metrics are produced to assess the effectiveness of the introduced algorithmic approach. The outcomes demonstrate that the new algorithm is superior to the current algorithmic technique. By offering a new security technique that uses machine learning models to separate cloud virtual channel attacks and can be compared to other models to demonstrate its validity, the research can be expanded.

REFERENCES

1. Indresh Gangwar, Poonam Rana, "Cloud Computing Overview: Services and Features", 2014, International Journal of Innovations & Advancement in Computer Science (IJIACS), Volume 3, Issue 1
2. Srinivasa Rao V, Nageswara Rao N K, E Kusuma Kumari, "Cloud Computing: An Overview", 2009, Journal of Theoretical and Applied Information Technology, Vol. 5, No. 2
3. Mahantesh N. Birje, Praveen S. Challagidad, R.H. Goudar, Manisha T. Tapale, "Cloud computing review: concepts, technology, challenges and security", 2017, Int. Journal of Cloud Computing, Vol. 6, No. 1
4. Deepa. B, Srigayathri.S and Visalakshi.S, "A Review on Cloud Computing", 2018, International Journal of Trend in Research and Development, Vol. 4. No. 1
5. Priyanshu Srivastava, Rizwan Khan, "A Review Paper on Cloud Computing", 2018, International Journals of Advanced Research in Computer Science and Software Engineering, Volume-8, Issue-6
6. Mohammad Oqail Ahmad, Dr.Rafiqul Zaman Khan, "The Cloud Computing: A Systematic Review", 2015, International Journal of Innovative Research in Computer and Communication Engineering, Vol. 3, Issue 5
7. Majid Mehmood, Kinza Sattar, Asif Hussain Khan3 and Mujahid Afzal, "Load Balancing Approach in Cloud Computing", 2015, Journal of Information Technology & Software Engineering, Vol. 5, No. 3
8. W. Z. Jiang and Z. Q. Sheng, "A New Task Scheduling Algorithm in Hybrid Cloud Environment," 2012 International Conference on Cloud and Service Computing, 2012, pp. 45-49

9. S. Bilgaiyan, S. Sagnika and M. Das, "Workflow scheduling in cloud computing environment using Cat Swarm Optimization," 2014 IEEE International Advance Computing Conference (IACC), 2014, pp. 680-685

10. N. Chaudhary and M. Kalra, "An improved Harmony Search algorithm with group technology model for scheduling workflows in cloud environment," 2017 4th IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics (UPCON), 2017, pp. 73-77

11. B. Wang, Y. Hou and M. Li, "QuickN: Practical and Secure Nearest Neighbor Search on Encrypted Large-Scale Data," in IEEE Transactions on Cloud Computing, vol. 1, no.16, pp 265-273, 2020

12. H. Zhu and Y. Jin, "Multi-Objective Evolutionary Federated Learning," in IEEE Transactions on Neural Networks and Learning Systems, vol. 31, no. 4, pp. 1310-1322, April 2020

13. Z. L. Jiang, H. Guo, Y. Pan, Y. Liu, X. Wang and J. Zhang, "Secure Neural Network in Federated Learning with Model Aggregation under Multiple Keys," 2021 8th IEEE International Conference on Cyber Security and Cloud Computing (CSCloud)/2021 7th IEEE International Conference on Edge Computing and Scalable Cloud (EdgeCom), 2021, pp. 47-52

14. T. Kong, L. Wang, D. Ma, K. Chen, Z. Xu and Y. Lu, "ConfigRand: A Moving Target Defense Framework against the Shared Kernel Information Leakages for Container-based Cloud," 2020 IEEE 22nd International Conference on High Performance Computing and Communications; IEEE 18th International Conference on Smart City; IEEE 6th International Conference on Data Science and Systems (HPCC/SmartCity/DSS), 2020, pp. 794-801

15. Z. Wu, K. Li, K. Li and J. Wang, "Fast Boolean Queries With Minimized Leakage for Encrypted Databases in Cloud Computing," in IEEE Access, vol. 7, pp. 49418-49431, 2019

16. B. H. K. Chen, P. Y. S. Cheung, P. Y. K. Cheung and Y. -K. Kwok, "CypherDB: A Novel Architecture for Outsourcing Secure Database Processing," in IEEE Transactions on Cloud Computing, vol. 6, no. 2, pp. 372-386, 1 April-June 2018

17. S. Gao, G. Piao, J. Zhu, X. Ma and J. Ma, "TrustAccess: A Trustworthy Secure Ciphertext-Policy and Attribute Hiding Access Control Scheme Based on Blockchain," in IEEE Transactions on Vehicular Technology, vol. 69, no. 6, pp. 5784-5798, June 2020

18. S. V. Usov and A. N. Mironenko, "Solving some optimization tasks of assigning roles within the framework of well-known estimates of relative damage from leakage of authority in a role-based access control model," 2019 Dynamics of Systems, Mechanisms and Machines (Dynamics), 2019, pp. 1-7

19. Y. Wang, Y. Ma, K. Xiang, Z. Liu and M. Li, "A Role-Based Access Control System Using Attribute-Based Encryption," 2018 International Conference on Big Data and Artificial Intelligence (BDAI), 2018, pp. 128-133

20. R. Ghazal, A. K. Malik, N. Qadeer, B. Raza, A. R. Shahid and H. Alquhayz, "Intelligent Role-Based Access Control Model and Framework Using Semantic Business Roles in Multi-Domain Environments," in IEEE Access, vol. 8, pp. 12253-12267, 2020

21. K. Soni and S. Kumar, "Comparison of RBAC and ABAC Security Models for Private Cloud," 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), 2019, pp. 584-587

22. K. Lee, "Comments on "Secure Data Sharing in Cloud Computing Using Revocable-Storage Identity-Based Encryption"," in IEEE Transactions on Cloud Computing, vol. 8, no. 4, pp. 1299-1300, 1 Oct.-Dec. 2020

23. Z. -Y. Liu, Y. -F. Tseng, R. Tso, Y. -C. Chen and M. Mambo, "Identity-Certifying Authority-Aided Identity-Based Searchable Encryption Framework in Cloud Systems," in IEEE Systems Journal, vol. 4, no. 7, pp. 1046-1052, 2021