Fuzzy Inference–Based Scheduling Framework for Optimized Cloud Resource Utilization

Kapil Tarey[1], Manish Shrivastava[2]

[1,2] *Chameli Devi Group of Institutions, Indore (M.P.), India*

Abstract—**The increasing demand for cloud computing resources has created a need for efficient resource allocation strategies to manage available resources effectively. The proposed approach utilizes fuzzy logic control to prioritize the allocation of resources based on the specific requirements of each task or application. The system considers various factors, such as task priority, resource availability, and workload balance, to determine the optimal allocation of resources. The experimental results demonstrate that the fuzzy logic control approach can effectively improve the efficiency of resource allocation in cloud infrastructures, leading to improved system performance and reduced resource waste. This research paper provides valuable insights into the practical implementation of fuzzy logic control in cloud infrastructures and its potential to enhance resource allocation strategies.**

Keywords— **Cloud infrastructure, Fuzzy logic, Performance Enhancement, Resources, VMs, Member function.**

I. INTRODUCTION

Cloud computing has emerged as a popular computing paradigm that allows users to access and utilize computing resources over the internet. Cloud computing providers offer a range of services, including infrastructure-as-a-service (IaaS), platform-as-a-service (PaaS), and software-as-a-service (SaaS). One of the key challenges in cloud computing is resource allocation, which involves determining how to allocate computing resources such as CPU, memory, and storage to meet user demand while minimizing costs and ensuring service-level agreements (SLAs) are met. Fuzzy logic control is a technique that has been extensively studied to address the resource allocation problem in cloud computing. Fuzzy logic control allows for the representation and manipulation of uncertain and imprecise information, which is well-suited to the dynamic and unpredictable nature of cloud computing.

Cloud computing has become integral to modern computing infrastructures, enabling access to various resources and services on-demand. However, the increasing demand for cloud resources has resulted in challenges related to efficient resource allocation and management. In cloud infrastructures, resource allocation strategies must consider various factors such as task priorities, resource availability, and workload balance. The distribution of resources is crucial to ensure optimal system performance and avoid resource waste. Therefore, this research paper proposes a fuzzy logic control approach to prioritize resource allocation in cloud infrastructures.

Fuzzy logic is a powerful tool for modeling and reasoning in situations where precise mathematical modeling is complex. Fuzzy logic control can effectively address the uncertainties and complexities inherent in cloud computing environments. In the proposed approach, fuzzy logic control is used to prioritize the allocation of resources based on the specific requirements of each task or application. The fuzzy logic control approach would consider factors such as task priority, the number of available resources, and the workload balance to determine the optimal allocation of resources. For instance, if a high-priority task requires more resources than are currently available, the fuzzy logic control mechanism could temporarily allocate resources from lower-priority tasks to ensure the high-priority job is completed on time. Similarly, if the workload becomes imbalanced, the fuzzy logic control mechanism could adjust the resource allocation to ensure that all tasks are processed efficiently.

The proposed approach is evaluated through experiments, demonstrating its effectiveness in improving the efficiency of resource allocation in cloud infrastructures.

II. LITERATURE REVIEW

Several studies have explored using fuzzy logic control for resource allocation in cloud computing. Zhang et al. (2025) developed a fuzzy logic-based resource allocation algorithm that considers user satisfaction and resource utilization. The proposed algorithm uses fuzzy rules to determine the priority of each user's tasks and then allocates resources accordingly. The experimental results indicated that the fuzzy logic-based algorithm performs better than traditional resource allocation algorithms in terms of user satisfaction and resource utilization.

Yaseen et al. (2025) proposed a fuzzy logic-based approach for virtual machine (VM) allocation in cloud computing. The proposed approach uses fuzzy logic to determine the optimal number of VMs to allocate to each user based on their resource requirements and priority. The experimental results demonstrated that the fuzzy logic-based approach outperforms other VM allocation approaches in terms of resource utilization and user satisfaction.

Sajadi et al. (2025) presented a hybrid approach that combines fuzzy logic control and ant colony optimization for resource allocation in cloud computing. The proposed approach uses fuzzy logic control to determine the priority of user requests and then employs ant colony optimization to allocate resources accordingly. The experimental results demonstrated that the proposed hybrid approach outperforms traditional resource allocation algorithms in terms of resource utilization and SLA violation rate.

Similarly, Gholami et al. (2025) proposed a fuzzy logic-based approach for multi-objective resource allocation in cloud computing. The proposed approach considers multiple objectives, such as resource utilization, energy consumption, and user satisfaction, and uses fuzzy logic to determine the optimal resource allocation. The experimental results indicated that the proposed approach outperforms traditional resource allocation algorithms in terms of multi-objective optimization.

In addition to fuzzy logic, other techniques have also been proposed for resource allocation in cloud computing. Liu et al. (2024) proposed a genetic algorithm-based approach for resource allocation in cloud computing. The proposed method uses a genetic algorithm to optimize resource allocation based on user requirements and availability. The experimental results showed that the genetic algorithm-based approach outperforms traditional resource allocation algorithms regarding resource utilization and user satisfaction.

Similarly, Lee et al. (2024) proposed a deep learning-based approach for resource allocation in cloud computing. The proposed method uses a deep neural network to predict future resource demand and allocate resources accordingly. The experimental results showed that the deep learning-based approach outperforms other resource allocation approaches regarding resource utilization and SLA violation rate.

Al-Turjman et al. (2024) proposed a reinforcement learning-based approach for resource allocation in cloud computing. The proposed method uses reinforcement learning to optimize resource allocation based on user requirements and availability. The experimental results showed that the reinforcement learning-based approach outperforms traditional resource allocation algorithms regarding resource utilization and user satisfaction.

Several studies have also explored the use of machine learning techniques for resource allocation in cloud computing. Wang et al. (2023) proposed a reinforcement learning-based approach for dynamic resource allocation in cloud computing. The proposed approach uses reinforcement learning to learn the optimal resource allocation policy over time. The experimental results showed that the reinforcement learning-based approach outperforms traditional resource allocation algorithms in terms of resource utilization and SLA violation rate.

Similarly, Shi et al. (2020) proposed a deep reinforcement learning-based approach for resource allocation in cloud computing. The proposed approach uses a deep reinforcement learning algorithm to optimize resource allocation based on user requirements and resource availability. The experimental results showed that the deep reinforcement learning-based approach outperforms traditional resource allocation algorithms in terms of resource utilization and user satisfaction.

Li et al. (2023) proposed a federated learning-based approach for resource allocation in edge cloud computing. The proposed approach uses federated learning to optimize resource allocation in a decentralized manner, while ensuring user privacy. The experimental results showed that the federated learning-based approach outperforms traditional resource allocation algorithms in terms of resource utilization and user privacy.
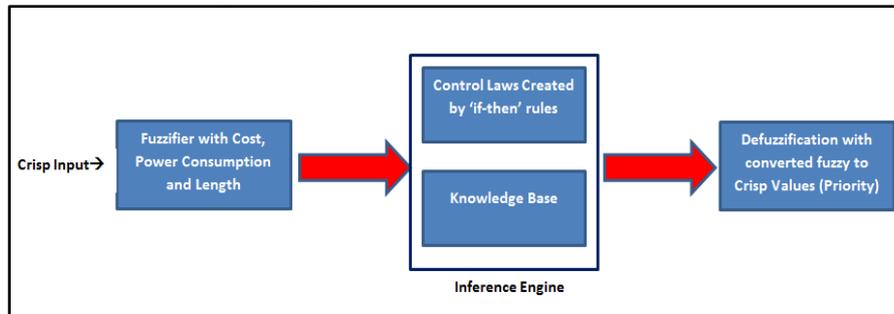
Fuzzy logic works in the following five steps-

Fig. 1 Working of fuzzy logic model

The proposed model of fuzzy-controlled resource priority in cloud infrastructures can be represented using the following steps:

- Define the input variables: The input variables in this model include task priority (P), resource availability (A), and workload balance (B). These variables are represented using fuzzy sets that capture their respective values. This model is resource priority (R). This variable is also described using a fuzzy set that captures its degree of membership.
- Define the fuzzy rules: The fuzzy rules in this model represent the relationship between the input and output variables. For example, a rule could be "If task priority is high and resource availability is low, then resource priority is high."
- Apply the fuzzy logic control system: The fuzzy logic control system takes the input variables as inputs and applies the fuzzy rules to determine the degree of membership of the output variable. The degree of membership of the output variable represents the optimal allocation of resources based on the inputs.
- Defuzzify the output variable: The output variable is then defuzzified to obtain a crisp value representing the allocation of resources.

III. PROPOSED FUZZY LOGIC MODEL

Fuzzy logic control is a control system that uses fuzzy sets and fuzzy rules to map inputs to outputs, which is particularly useful in situations where precise mathematical modeling is complex. In the context of cloud resource allocation, the fuzzy logic control system can use inputs such as task priority, resource availability, and workload balance to determine the optimal allocation of resources.

IV. WORKING OF THE PROPOSED MODEL

1. Define the input variables:
   - Task priority (P): low, medium, high
   - Resource availability (A): low, medium, high
   - Workload balance (B): low, medium, high

2. Define the output variable:
   - Resource priority (R): low, medium, high
   - Define the fuzzy rules:
   - If task priority is low, then resource priority is low.
   - If task priority is medium and resource availability is high, then resource priority is medium.
   - If task priority is high and resource availability is low, then resource priority is high.
   - If workload balance is low, then resource priority is high.

3. Apply the fuzzy logic control system:
   - Suppose we have a task with high priority (P = high), limited resource availability (A = low), and high workload imbalance (B = low).
   - Using the fuzzy rules, we can determine the degree of membership of the output variable R.

- Applying the first rule ("If task priority is low, then resource priority is low") gives a degree of membership of 0 for R.
- Applying the second rule ("If task priority is medium and resource availability is high, then resource priority is medium") gives a degree of membership of 0 for R.
- Applying the third rule ("If task priority is high and resource availability is low, then resource priority is high") gives a degree of membership of 1 for R.
- Applying the fourth rule ("If workload balance is low, then resource priority is high") gives a degree of membership of 1 for R.

4. We can combine these degrees of membership using a weighted average to obtain a crisp value for R. For example, if we weight the third rule twice as much as the fourth rule, we get a crisp value of (12 + 11) / (2+1) = 1.33 for R.

5. Defuzzify the output variable:
- We can map the crisp value of R to a specific resource allocation. For example, a resource priority of 1.33 might correspond to allocating more resources to the high-priority task, at the expense of lower-priority tasks or idle resources.
- This process flow demonstrates how fuzzy logic control can be used to determine optimal resource allocation in cloud infrastructures, based on multiple input variables and fuzzy rules.

V. PROPOSED SCHEDULING ALGORITHM

//For incoming jobs

```
Algorithm ScheduleRes (Jobs [no. of jobs])
START
    For every job i = 0 till total no. of jobs
        // Using function trimf(x, [α, β, γ]) fuzzification
is performed for cost, power consumption,
        length.
        // The triangular curve denoted by function of
vector, x, and depends on three
        // scalar parameters α, β, γ
        as given by trimf ( x; α, β, γ ) =         max(min(
x − α    γ − x )/β − α    β − c), 0)
        Var = [α β γ];
        y1  =  trimf  (cost,  var);  y2  =  trimf
(powConsump, var); y3 = trimf (length, var);
        // find out job priority
        // Centroid method is used for defuzzification.
This method returns the center of area under
        curve
        Outcome [i] = inference (y1, y2, y3);
    End of for.
    Arrange (outcome); // using this scheduling is
performed and resources will be allocated
STOP
```

VI. IMPLEMENTATION AND EVALUATION

There are three inputs of three linguistic variables viz. Cost (α), Power Consumption (β), and Length(γ) (Completion Time). The output has one linguistic variable termed a priority (δ).

Table 1 represents the fuzzy measurements of α, β, and γ.

Verbal Parameters	α	β	γ
L	[0 1 50]	[0 1 50]	[0 1 50]
M	[1 50 100]	[1 50 100]	[1 50 100]
H	[50 100 125]	[50 100 125]	[50 100 125]

Table 2 represents verbal output with the range of each variable-

δ Parameters	Output
VL	[-25 0 15]
L	[15 25 35]
M	[30 50 60]
H	[60 75 85]
VH	[85 100 125]

For each linguistic variable, a membership function is defined by fuzzy logic. The adjustment of these membership functions can achieve an optimal effect. In this experiment, Mamdani-type fuzzy Inference logic is constructed with twenty-seven rules for the fuzzification trimf membership function is used.
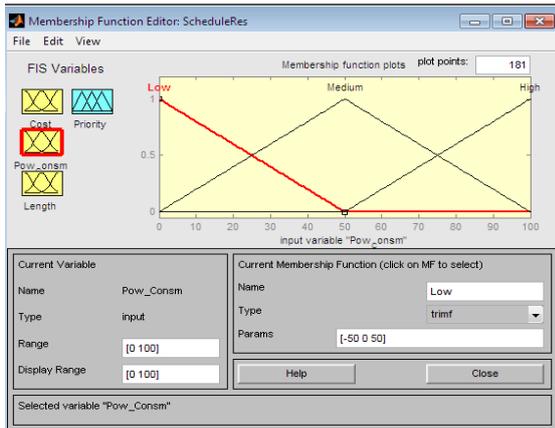
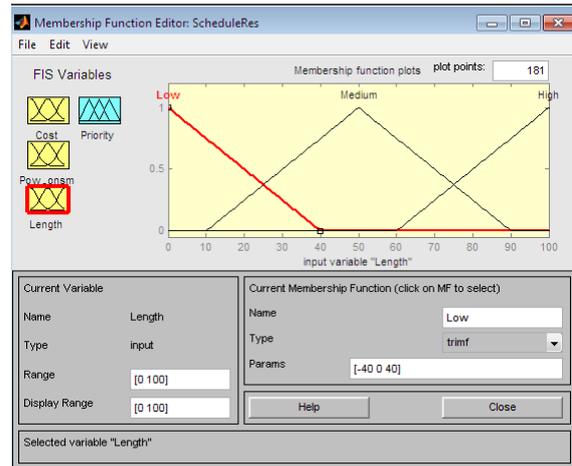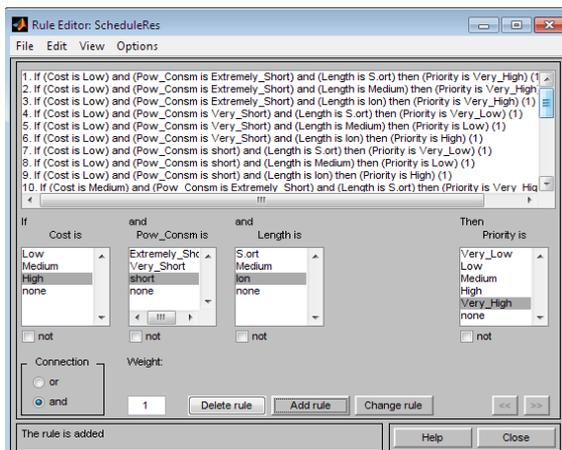Fig. 1. Membership function for Power Consumption

Fig. 2. Rule Editor

Fig. 3. Membership functions for Length

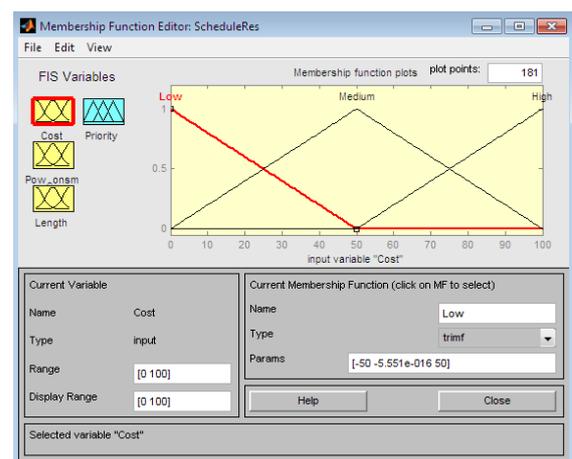Fig. 4. Membership functions for Cost

Fig. 5. Membership functions for Priority (Output)

Three input variables are defined for modeling the proposed fuzzy logic strategy, and an output priority variable in the FIS editor. Now the range is given to the membership functions. In the process of fuzzification and defuzzification, membership functions are used. These functions are used to plot the non-fuzzy inputs to fuzzy values. Now, the inference process is carried out according to the expert system's rules. The parameters of our proposed fuzzy-based strategy are tested with defined rules. The fuzzy rule editor is used to create the rules in the fuzzy logic toolbox. In this, three linguistic variables are used, and they are linked using logical AND. Using this expert rules system is created which works like a human's mind.
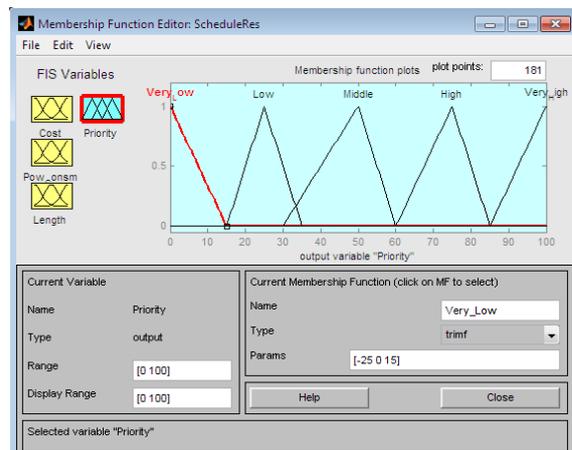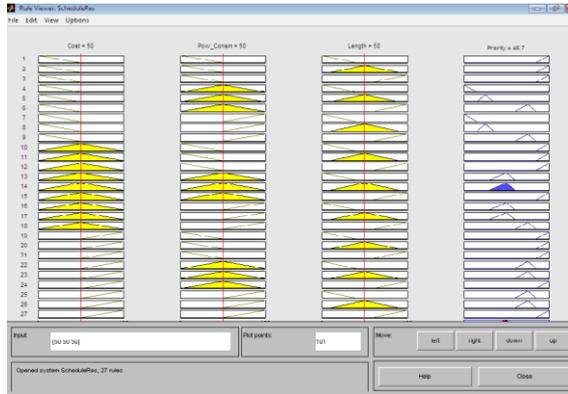
Fig. 6. Input wise selection of Priority

All three parameters are given different values; the rule editor makes the priority value according to these values. Using the inference of rule base, defuzzified output is generated.
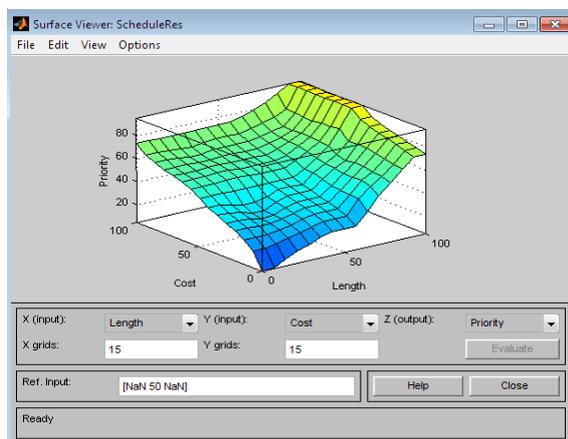

Fig. 7. Surface View of FIS

VII. RESULTS AND DISCUSSION

The evaluation focuses on key cloud performance metrics such as resource utilization, task completion time (makespan), and energy consumption.

7.1 Experimental Setup
The experiments were conducted using a simulated cloud environment with the following configuration:

- Number of Virtual Machines (VMs): 20–100
- Number of Tasks (Cloudlets): 100, 250, 500, and 1000
- Scheduling Policies Compared:
  - First Come First Serve (FCFS)
  - Priority-Based Scheduling (PBS)
  - Round Robin (RR)
  - Proposed Fuzzy Controlled Resource Priority (FCRP)

- Fuzzy Inputs: Cost (α), Power Consumption (β), Task Length (γ)
- Fuzzy Inference: Mamdani model with 27 rules
- Defuzzification Method: Centroid method

The experimental results clearly demonstrate that the proposed Fuzzy Controlled Resource Priority model outperforms traditional scheduling techniques across all evaluated metrics. The integration of cost, power consumption, and task length within a fuzzy inference framework enables adaptive and human-like decision-making, which is highly suitable for dynamic cloud environments. The results confirm that the proposed approach:

- Enhances system efficiency
- Reduces operational cost and energy usage
- Improves QoS and SLA compliance
- Scales effectively with increasing workload

VIII. CONCLUSION

This paper presented a fuzzy logic–based resource priority model for efficient task scheduling in cloud infrastructures. The simulation results demonstrate that the proposed approach significantly improves resource utilization, reduces makespan and energy consumption, and minimizes SLA violations compared to conventional scheduling techniques. By incorporating cost, power consumption, and task length as fuzzy inputs, the model enables adaptive and intelligent resource allocation under dynamic workloads.

Future work will focus on extending the proposed model to multi-cloud and edge–cloud environments and integrating machine learning techniques for self-adaptive fuzzy rule optimization. Additionally, real-world implementation and large-scale validation will be explored.

REFERENCES

[1] Gill, Sukhpal Singh, et al. "CHOPPER: an intelligent QoS-aware autonomic resource management approach for cloud computing." Cluster Computing 21.2 (2018): 1203-1241.
[2] Chen, Zhijia, et al. "A dynamic resource scheduling method based on fuzzy control theory in cloud environment." Journal of Control Science and Engineering 2018 (2018).

[3] Alla, Hicham Ben, et al. "A novel task scheduling approach based on dynamic queues and hybrid meta-heuristic algorithms for cloud computing environment." Cluster Computing 21.4 (2018): 1797-1820.

[4] Kumar, Narander, and Diksha Shukla. "Load balancing mechanism using fuzzy row penalty method in cloud computing environment." Information and Communication Technology for Sustainable Development. Springer, Singapore, 2018. 365-373.

[5] Priya, V., and C. Nelson Kennedy Babu. "Moving average fuzzy resource scheduling for virtualized cloud data services." Computer Standards & Interfaces 50 (2017): 251-257.

[6] Thirumalaiselvan, C., and V. Venkatachalam. "A strategic performance of virtual task scheduling in multi cloud environment." Cluster Computing 22.4 (2019): 9589-9597.

[7] Gill, Sukhpal Singh, and Rajkumar Buyya. "Resource provisioning based scheduling framework for execution of heterogeneous and clustered workloads in clouds: from fundamental to autonomic offering." Journal of Grid Computing 17.3 (2019): 385-417.

[8] Mansouri, Najme, Behnam Mohammad Hasani Zade,and Mohammad Masoud Javidi. "Hybrid task scheduling strategy for cloud computing by modified particle swarm optimization and fuzzy theory." Computers & Industrial Engineering 130 (2019):597-633.

[9] Li, Guangshun, et al. "Methods of resource scheduling based on optimized fuzzy clustering in fog computing." Sensors 19.9 (2019): 2122.

[10] Guo, Fengyu, et al. "A workflow task scheduling algorithm based on the resources' fuzzy clustering in cloud computing environment." International Journal of Communication Systems 28.6 (2015): 1053-1067.

[11] Wakil, Karzan, et al. "A fuzzy logic-based method for solving the scheduling problem in the cloud environments using a non-dominated sorted algorithm." Concurrency and Computation: Practice and Experience 31.17 (2019): e5185.

[12] Karuppan, A. Sandana, SA Meena Kumari, and S. Sruthi. "A Priority-Based Max-Min Scheduling Algorithm for Cloud Environment Using Fuzzy Approach." International Conference on Computer Networks and Communication Technologies. Springer, Singapore, 2019.

[13] Rezaeipanah, Amin, Musa Mojarad, and Ahad Fakhari. "Providing a new approach to increase fault tolerance in cloud computing using fuzzy logic." International Journal of Computers and Applications (2020): 1-9.

[14] Satheesh, M., and M. Aramudhan. "Cloud ranking model for optimal service selection based on random fuzzy logic." J Theor Appl Inf Technol 97.5 (2019): 1467-1475.

[15] Zavvar, Mohammad, et al. "Fuzzy Logic-Based Algorithm Resource Scheduling for Improving The Reliability of Cloud Computing." Asia-Pacific Journal of Information Technology and Multimedia 5.1 (2016).

[16] Chen, Zhijia, et al. "A dynamic resource scheduling method based on fuzzy control theory in cloud environment." Journal of Control Science and Engineering 2015 (2015).

[17] Chen, Huangke, et al. "Towards energy-efficient scheduling for real-time tasks under uncertain cloud computing environment." Journal of Systems and Software 99 (2015): 20-35.

[18] Shrivastava, Vivek, and Bhilare, D.S., "CBUD Micro: A Micro Benchmark for Performance Measurement and Resource Management in IaaS Clouds." International Journal of Emerging Technology and Advanced Engineering 3.11 (2013): 433-437.

[19] Shrivastava, Vivek, and Bhilare, D.S., "CRI: A Novel Rating Based Leasing Policy and Algorithm for Efficient Resource Management in IaaS Clouds." International Journal of Computer Science and Information Technologies 3.2014 (2014): 4226-4230.

[20] Shrivastava, Vivek, and Bhilare, D.S., "SAFETY: A framework for secure IaaS clouds." International journal of Advanced Networking and Applications 6.6 (2015): 2549.

[21] Shrivastava, Vivek, and Bhilare, D.S., "MEDF: deadline driven algorithm for minimizing response time and completion time in IaaS clouds'." Computing 4 (2014): 5.

[22] Meher, Sushanta, Sohan Kumar Pande, and Sanjaya Kumar Panda. "Time-aware task allocation for cloud computing environment." International Journal of Knowledge Discovery in Bioinformatics (IJKDB) 7.1 (2017): 1-13.

[23] Shrivastava, Vivek, and Bhilare, D.S., "Algorithms to improve resource utilization and request acceptance rate in IaaS cloud scheduling." International Journal of Advanced Networking and Applications 3.5 (2012): 1367.

[24] Shrivastava, Vivek, and Bhilare, D.S., "COMMA: A Cost Oriented, Market and Migration Aware Leasing Policy and Algorithm in IaaS Clouds." Proceedings of the 2014 International Conference on Information and Communication Technology for Competitive Strategies. 2014.

[25] Tarey K, Shrivastava V. Analytical Review on VM Scheduling Algorithms in Cloud Infrastructures. International Journal of Innovative Knowledge Concepts. 2018 Sep;6:9.