MACHINE LEARNING METHODS FOR REFINING SLA BASED ADMISSION CONTROL AND RESOURCE ALLOCATION IN CLOUD COMPUTING

M.Sri Lakshmi
Assistant professor, Dept. of CSE,
G.Pullaiah College of Engineering and Technology,
Kurnool,India

Dr.S.Prem Kumar
Professor, Dept. of CSE,
G.Pullaiah College of Engineering and Technology,
Kurnool,India

M.Janardhan
Associate professor,
Dept. of CSE,
G.Pullaiah College of Engineering and Technology,
Kurnool,India

K.Gayathri
Assistant professor,
Dept. of CSE,
G.Pullaiah college of Engineering and Technology,
Kurnool,India

Abstract: Cloud computing is the most advanced technology in the real world environment and provides flexible and convenient possibilities for users to utilize available services. Resource provisioning to the satisfaction of user requirements becomes the most challenging task in the heterogeneous cloud environment. Proper admission control algorithms need to be proposed for better resource provisioning with improved user satisfaction level. In this research, Knowledge-based Service Level Agreement (SLA) aware admission controlled scheduling and resource allocation are proposed which makes use of machine learning algorithms namely Support Vector Machine (SVM) and Artificial Neural Network (ANN) for better admission control. It seeks to study the knowledge of resource status information by using machine learning algorithms in the training phase. Based on these strategies, admission control would be done in the testing phase which would lead to efficient and better resource provisioning. As our proposed work we mentioned Position Balanced Parallel Particle Swarm Optimization (PBPPSO) is utilized for optimal scheduling and resource allocation which can handle large volumes of tasks in an optimal manner.

Keywords: Cloud Computing, Position Balanced Parallel Particle Swarm Optimization, Service Level Agreement.

1. INTRODUCTION

Resource allocation is the process of selecting and providing the most suitable resources that satisfy user requirements. Resource allocation is the most complex task in the cloud computing environment where the selection of resources with the satisfaction of Quality of Service (QoS) would be more challenging of the tasks. The resources need to be utilized well for achieving higher profit for the cloud services providers. A lot of research work has been conducted on optimal resource allocation [1].The users can utilize the resources that are allocated by the cloud computing service providers through virtualization environment. Virtualization enables service providers to handle multiple user's requests simultaneously. Load level of virtual machines needs to be balanced for better. Resource utilization so that the tasks can be completed in a competent manner.

Resource allocation is divided into three main types based on the allocation methodology [2]. They are

- Static Resource Allocation
- Dynamic Resource Allocation
- Users Self Allocation

One of the main goals which need to be satisfied by different Resource allocation techniques are Service Level Agreement (SLA). SLA is defined as the contract made between the cloud service providers and the cloud users regarding some QoS constraints for completing the task execution submitted by the cloud users. These resource allocation techniques should ensure the SLA for the assured delivery of resources and complete execution of tasks submitted by the user. A brief explanation of the static and dynamic resource allocation is given in the following subsection.

This paper organized as chapter 2: Literature survey, Chapter-III .Problem Definition, Chapter-4: Objectives, Chapter-5 Contribution of the work. Chapter-6 Proposed Framework, Chapter 7. Comparison of SVM with ANN, Chapter 8.Results and Chapter 9. Conclusion of the work.

2. LITERATURE SURVEY

ADMISSION CONTROL IN CLOUD COMPUTING

Mardente et al. [3] designed a novel entity called domain controller involving cloud tasks and cloud service providers to perform the admission control process. Domain controller will perform admission control tasks by analyzing the factor called the traffic flow. The domain controller approach utilizes traffic flow detection algorithms to detect the level of traffic flow in resources. It will prevent the task from submission to the cloud resource that has more traffic flow where there might be a possibility of task execution failure

Rahul Garg et.al [4] the author has been presented SLA centered framework for QoS distribution and dynamic capacity allocation. This methodology works based on two types of pricing strategies. The cloud service provider would maintain a fixed price for the long-term utilized resources and the dynamic pricing for the short terms resources which are

reserved dynamically. This work also introduces the third tier, which will fix the incentives and penalties in run time. The overall research of this work provides a continuous allocation of resources to the cloud user with guaranteed SLA.☐

Vipul Mathur et al.[5] proposed an adaptive admission control scheme to provide better scheduling performance for web applications with variable capacity. The approach adopts dynamically by adjusting the load to compensate for the variation in the system capacity. The scheme is implemented as a proxy server involving the clients and front-end web servers. The main purpose is to stabilize the changes in response time and throughput, while also avoiding overloads that occur as a result of a reduction in available system capacity.

Jaideep Dhok et al.[6] the author has been disclose the OACA based on knowledge methods for the case MapReduce as a service. The admission control algorithm allows the user to buffer the MapReduce jobs in the form of on-demand services. In this approach, a learning algorithm is used to model MapReduce jobs in a ready to use manner which effectively makes the MapReduce as SaaS. The admission control algorithm is based on machine learning algorithm for the prediction of the job admission. Resource sharing and multiplexing the job execution as an alternative to focusing on maximizing user and service provider utility improves the efficiency of the job scheduling approach.☐

MACHINE LEARNING BASED RESOURCE ALLOCATION

Gaurav Pandey et al.[7] introduced novel active Support Vector (SV) learning algorithms that can adapt the combinations of random and query learning. These mechanisms can perform scheduling with the satisfaction of users QoS constraints. The proposed algorithms provide better results regarding achieving fast and robust convergence. Random SVM would assume that all the predicted patterns are independent of each other by choosing the correct predictable answer. In query SVM, the output would be predicted by finding the solution that is most near convergence to the hyper plane which depicts the QoS factors. This approach can reach a most optimal solution by following the convergence solution.☐

Josep L Berral et al.[8] proposed the aware energy scheduling using machine learning algorithms to provide a flexible environment for the service providers to scale down their energy consumption. The energy consumption is often related to factors such QoS parameters like power consumption levels, CPU loads, and SLA timings and so on. These need to be estimated well for the better prediction of the resources regarding achieving 22 enhanced scales down control of energy efficiency. The overall research of this work was evaluated in the practical heterogeneous cloud environment by considering the QoS parameters. This research work leads to better resource handling capability considering scaling down or scaling up of the resources.☐

OPTIMAL RESOURCE PROVISION

Sivadon Chaisiri et al.[9] introduced the optimal cloud resource provisioning (OCRP) algorithm which focuses on the allocation of the resources to the cloud consumers considering the reduction of overall cost of execution. The presented approach could be utilized for the long period for efficient and continuous allocation of resources such as the multi-resource stage allocation. The overall research of this work is divided into the four stages in the quarter plan and two stages of the yearly plan. This approach attempts to allocate the resources based on the demand level which would vary in different levels regarding the reduced cost.Manufacturers of recent multiprocessors have to adopt caches for cloud computing, particularly in the multitenant, multi-process, multithreading and dynamic environment.

Marjan Gusev et al. [10] exhibited an investigation of the single and multi-tenant clouds introduced on a solitary chip multi-core multiprocessor with the distinctive portion of resources to the tenants. The computational concentrated and memory requesting calculations are utilized to break down the execution of the diverse workload and variable store prerequisites. The majority of the examinations utilize similar resources in various collaborations. Though the virtualization and clouds causeˊ software overhead, this approach provides better performance than the traditional methods. ☐

Giridas et al. [11] the author has been influencing known to an optimal resource to provision practice (ORPP) for aide the green cloud computing. The technique grows the provision of resources of green cloud computing settings, by utilizing both the data transfer capacity and handling skill.. Then the resources are allocated to each service request, and also it is returned to it on a timely basis. From the simulation results, ORAT method diminishes the request loss possibility and reduces the total resource mandatory judged against with the predictable allocation schemes. Over and finished with the optimal resource provision, the resources for the obligations are relegated to cloud computing condition by disposing of e-waste and make IT as Green IT.. Though the approach supports improved green cloud, the performance is not sufficient.☐

GROUPING RESPONSIVE RESOURCE PROVISION

Bo An et al. [12] improved the resource allocation process in the multi-agent system with incomplete information. This was done by augmenting the resource allocation scheme with the optimization strategy which attempted to achieve better resource allocation strategy by resubmitting bid details to the resource providers. This resubmission of bid values optimizes the dealing stage, and thus the optimal resource allocation can thus be achieved.☐

Kevin Power et al.[13] introduced the novel mechanism for performing resource allocation process for the real-time services by grouping the resources. This grouping of resources is done to satisfy the heterogeneous resource requisition of real-time applications. This work gathers the user requirements which would be grouped based on the QoS constraints. The group of users with similar QoS constraints would be assigned to the shared resource that satisfies QoS constraints. After this initial allocation, data flow rate will be calculated regarding QoS constraints based on which resource allocation would be redone to achieve it. This group based resource allocation strategy achieves promised optimal resource allocation

scenario with the improved satisfaction of user QoS constraints. □

MAP REDUCE AWARE TASK SCHEDULING

Thirumala Rao & Reddy [14] researched scheduling data intensive aware application which would be more complex on the map-reduce environment. This approach is used for an improved handling of the large volume of user requests in the Hadoop environment using the map-reduce component. These works make use of two components for achieving the better scheduling results. These are resource predictor and resource reconfiguration. Resource predictor would identify the available resource information periodically. Resource configuration will reallocate the resources by adjusting the available resources for the remaining resources in an optimized manner. □

3. PROBLEM DEFINITION

The various research issues found in previous research methodologies are given below:

- It is very challenging to identify the proper resource for the user submitted task by making assumptions about the execution time in SLA based admission control. The assumption of execution time might lead to the wrong prediction of profit which would violate the proper allocation of resources. □

- Optimal resource allocation during the arrival of the large volume of tasks cannot be done efficiently using machine learning approaches. □

- Number of tasks might cause the generation of more number of rules based on best fitness value in the optimization approaches. Number of rules will lead to more computation overheads. □

- Data transfer cost would be more in the case of arrival of the large volume of user requests with different input data requirement that are located in different places. The prioritized users are not given importance in previous methodologies. □

4. OBJECTIVES

The present research work focuses on optimal resource allocation for the cloud consumer jobs with consideration of the multiple QoS constraints with reduced execution time. The major objectives of the proposed research work are:

- To predict the near future resource information accurately by learning historical information about resources using machine learning approaches (Support Vector Machine and Artificial Neural Network) to achieve proper resource allocation with improved user satisfaction level. □
- To optimize the resource allocation process by utilizing the optimization algorithm instead of machine learning algorithms to support huge volume

of tasks and to avoid incorrect indelicate parameters value which may lead to wrong predictions.

- To eliminate the computation overheads that occur due to the generation of large volume of rules for more number of arrived tasks by introducing the NMF based clustering approach.

- To provide priority for the users based on their category and to reduce the data transfer cost by selecting the shortest path from which data can be retrieved for the execution.

5. CONTRIBUTION OF THE WORK

The first contribution is to propose a machine learning based approach for SLA aware resource allocation with admission control. This research considers two machine learning approaches for SLA aware resource allocation, namely Support vector machine (SVM) classification approach, and the Artificial Neural Network (ANN) classification approach. In this approach, QoS constraints like profit and the make span are considered for improved user satisfaction level.

The second contribution of this paper is the Position Balanced Parallel Particle Swarm Optimization (PB-PPSO) method utilized for optimizing the resource allocation process. This approach extracts rules from the training set and then finds the optimal rule based on the best fitness values. By using these rules, optimal resource allocation with the satisfaction of profit and makespan parameters is ensured for the current allocation.

The third contribution of this paper is NMF-based optimal clustering and PB-PPSO based resource Prediction method (NMF-PB-PPSO). This approach is used to efficiently handle large volumes of tasks with reduced time by clustering similar kind of tasks and resources together. Once clustering, optimal resource allocation for those clusters is done by using the Position Balanced Parallel Particle Swarm Optimization approach(PPSO). □

6. PROPOSED FRAMEWORK

6.1 Machine Learning Approaches in Refining Service Level Agreement-Based Admission Control On behalf of a SaaS Provider in Cloud

Software-as-a-Service provider is popular in cloud computing environment which offers a set of reliable services to the cloud consumers regarding their request. To do so, SaaS service providers lease the required resources from the IaaS service providers to achieve their client requirements. Renting resources from the external service providers such as IaaS service providers might increase the administration and maintenance costs of the internal service providers. The rental system of the cloud collects the required resources from the external service provider which would also cause the service quality degradation. To overcome these issues, this proposed research methodology seeks to use, machine learning approaches for ensuring the SLA based admission control. Two machine learning approaches that are proposed in this research work for reaching the SLA based admission control are Support Vector Machine (SVM) approach and Artificial Neural Network (ANN) approach. Both of the proposed

approaches attempt to control the admission control process by allocating the tasks to such resources that can provide more Return on Investment (ROI). This is done by learning the profit and the non-profit parameters of the cloud resources and the tasks in an enhanced manner.☐

SLA satisfaction is a very important factor in cloud computing environment and must be considered for better satisfaction of user requirements through successful task execution. There might be the occurrence of more problems while attempting to perform admission control considering the SLA factors. These are☐

- Accepting latest requests without impacting already accepted requests.

- Mapping distinctive client demands with various QoS parameters to VMs.

- Deciding upon whether the latest request should be assigned to available VM or new VM must be initiated.

The present research work seeks to solve the above-mentioned issues by performing the admission control process by using machine learning methodologies such as SVM and ANN. The aim of these methods is to maximize the number of efficient user request placement on the VMs. Different customer QoS requirements and heterogeneity of infrastructure are also considered for attaining improved user satisfaction level☐

6.2 ADMISSION CONTROL IN CLOUD COMPUTING

SLA violation is the critical issue which might occur while renting the resources from the external service providers for internal deployment. This problem might create more impact on the software services that are allocated to cloud users regarding increased administration and maintenance cost. This problem needs to be resolved effectively for the better resource allocation to the cloud users. Hence the better admission control mechanism needs to be implemented between the SaaS providers, and the IaaS providers which can The proposed research methodology of this work makes use of two machine learning algorithms for efficient admission control with the satisfaction of various QoS requirements by learning the machine status information. The proposed research attempts to improve the satisfaction level of both cloud users and cloud service providers by considering SLA parameters of both. The admission control algorithm proposed in this work will analyze the various user QoS requirement and learned knowledge of machine status and then will find the suitable resources for better allocation. The scheduling would be performed in view of the admission control algorithms choice.

6.3 ADMISSION CONTROL WITH QoS FACTORS CONSIDERATION

Quality of Service plays vital role in the cloud computing environment where every user needs to complete their tasks as they want. In cloud computing condition, both cloud clients and cloud specialist co-ops have their necessities to satisfy goals, for example, benefit notoriety et cetera. In the proposed research methodology, QoS constraints of both service providers and cloud users are considered for admission control. SLA agreement is defined as the agreement made between cloud service providers and cloud users regarding sets of constraints. SLA agreement would be made, once both the cloud clients and the cloud specialist organizations concur with each other in light of their QoS limitations. This SLA agreement should be maintained until completion of all the tasks that are submitted by the users. QoS constraints considered for SLA agreement which is made regarding both cloud service provider and the cloud user for achieving better admission control is given in the following sub sections.☐

Service Level Agreement Constraints of Cloud Users

The admission control and scheduling algorithms are utilized to permit or reject a demand. A formal agreement-SLA (U) is signed between the user and provider only when the request is accepted, and the QoS requirements can be guaranteed with the following properties:☐

Deadline: Maximal time client might want to sit tight on the outcome.
Budget: Amount, client, will pay for the asked for administrations.
Penalty Rate Ratio: Amount gave for buyer's remuneration when the SaaS supplier misses the due date.
Input File Size: The measure of client's info record.
Request Length: Amount of Millions of Instructions (MI) required to be executed to serve the specific client's demand

In the proposed research work namely machine learning based admission control mechanism, the main objective considered is to improve return on investment (ROI). ROI is the most common profitability ratio. ROI estimates the amount of return on an investment relative to the investment's cost. ROI can be calculated by dividing the benefit of investment by the cost of the investment whose result is expressed in percentage or a ratio. The ROI of the service providers and as well as cloud users should be more of a better satisfaction level. Linlin Wu et al.[15] proposed mathematical equation for ROI calculation is given as follows:☐

6.4 MACHINE LEARNING BASED ADMISSION CONTROL

Machine learning techniques are most often used in the real world applications and are more suitable for the dynamic cloud environment. They can adapt to the dynamic nature of the resources and tasks flexibly. In this research work, two machines learning algorithms are proposed for better admission control by learning the resource status information effectively. Those algorithms are☐

- Support Vector Machine
- Artificial Neural Network

Both of the algorithms are evaluated in this work for admission control process, and outcomes of these algorithms are finally compared with each other to predict a better mechanism that can produce improved results regarding increased ROI. The working procedures of these methodologies are discussed in detail in the following sub sections.☐

Support Vector Machine Based Admission Control
Support vector machine is an administered learning technique which is utilized to investigate information and perceive patterns.It is also used for regression analysis and classification. Generally, SVM employs two possible classes for each of the input that takes a set of input data and predicts the admission. The conceivable classes frame the contribution by making the SVM a non-probabilistic paired direct classifier.

Artificial Neural Network Based Admission Control

Artificial Neural Networks (ANN) comprises of frameworks that are purposely worked to make utilization of some authoritative esteems looking like those of the human mind. They speak to the promising new making of data preparing frameworks. Neural Networks are great at assignments, for example, pattern coordinating and arrangement, optimization and information bunching. They have countless interconnected preparing components called neurons, which more often than not work in parallel and are sorted out in standard structures. The aggregate conduct of a NN, similar to a human cerebrum, uncovers the capacity to learn, review and sum up from preparing patterns or information. NNs are characterized by

i) Pattern of interconnection between neurons
ii) Learning algorithm
iii) Activation function.

In an NN, each neuron is connected to the other neuron using directed connected link and with an associated weight. Every ha an interior gaze called as its movement level. In light of the flag stream course, they are divided as bolster forward networks and criticism networks.

The add-on is the basic three components of the neuronal model:
i) A gathering of interfacing joins called neurotransmitters; each of the neurotransmitters is portrayed by weight or quality of its own. A flag xj at the contribution of neurotransmitters j associated with the neuron k is duplicated by weight wkj.
ii) An adder that performs the summation of the input signals.
iii) An activation function to control the amplitude of the neuron.

6.5 POSITION BALANCED PARALLEL PARTICLE SWARM OPTIMIZATION METHOD FOR RESOURCE ALLOCATION

Scheduling and allocation of resources play a critical role in the cloud environment which might violate the satisfaction level of both cloud users and the cloud service providers. Scheduling would be the most critical task in case of arrival of more number of tasks dynamically. This needs to be optimized to get a better solution of task resource scheduling. This research work attempts to handle number of incoming tasks that are arriving dynamically by scheduling them in an optimal manner. In this proposed inquire about, Position balanced Parallel Particle Swarm Optimization (PB-PPSO) conspire is produced for proficient designation of cloud assets. In this method, Particle Swarm Optimization (PSO) an unsupervised learning algorithm is used for finding the optimized resources for the group of tasks. In the optimization algorithm, every

particle has the gathering of errands and gathering of assets that starts with a self-assertive statement of particle's area and speed. Each particle in the swarm conduct has two particulars: an area which speaks to the proposed area and a speed which implies the speed of moving particle. The particle in the swarm consults over the whole pursuit space and remembers the best area found. The correspondence happens between the particles with the goal that they alter their areas and speeds in light of arrangements found by others. The area of the particle is scored by the wellness. The wellness is figured into the goal capacities. The principle target work is to give errand assignments that will achieve least makespan and least cost for the clients.

Based on the fitness value, the particle is quantified as a good solution. Among the execution of the PSO algorithm, the best wellness esteem is considered as the individual best wellness esteem. Looking at the whole particles in the swarm, the best wellness esteem is called worldwide wellness esteem. Besides, the weighted mean esteem is figured for the nearby best and worldwide best areas for decreasing the computational time. Every time the particle location and velocity is updated to move towards the local best solution. At long last, the worldwide best is recognized for the whole swarm. Thus, the streamlined arrangement is distinguished which has a gathering of assignments and gathering assets at least makespan and least cost. As per the improved arrangement, the standards are created in the preparation organize. In the training process, there are two class labels profit and loss. The label "loss" is assigned for the littler wellness estimation of the particle toward the finish of the emphasis. (i.e., errands running in these gathering of assets take high makespan and high cost). The name "benefit" is doled out for the more prominent wellness estimation of the particle toward the finish of the emphasis. The training phase is thus completed. If a new user gives the request, the rules are learned, and the resource with least make span and low cost is assigned.

ALGORITHM 1: POSITION BALANCED PARALLEL PARTICLE SWARM OPTIMIZATION (PB-PPSO) ALGORITHM

Input: Training samples and class labels
Output: Resource allocation

1. Initialize N number of particles with set of tasks and allocate the resources randomly, a position of particle is denoted by Xi and velocity is denoted as Vi .
2. *pbest* represents the best well-known position of particle i and gbest signifies the best position of the entire swarm
3.Particle position is initialized as Xi
4. For every particle i=1, 2… N

5. Compute the fitness value for each particle
6. // Fitness computation
7. Fitness Min = (Makespan, total cost)
8. If the fitness value is higher than the pbest
9. Set the present value as the new pBest
10. Until a termination criterion is met
11.Select the particle with best fitness value of all particles as the gbest
12. Update particle position and velocity
13. (t 1) x (t) v (t 1) i ii x += + +

14. until some stopping condition is met
15. Generate the rules and assign class labels
16. Particles which has less fitness are assigned to "loss" at every iteration
17. Particles which has high fitness are assigned to "profit" at every iteration
18.// Testing Process
19. If a new task is submitted, learn the rules
20. Assign the resources for the tasks

7. COMPARISON OF SVM WITH ANN

The performance of the SVM methodology with the ANN methodology is done to find the better algorithm that can perform admission control in an effective manner.

Table 1. Comparison analysis of SVM and ANN

S.No	Parameter	SVM	ANN
1	Efficiency	Depends on initial selection of parameters	More efficient than SVM
2	Flexibility	Inflexible to the dynamic cloud environment	More dynamic in nature
3	Parametric Model	Static	Can adapt to the environment dynamically
4	Input format	Must be unique	Not restricted to input format
5	Online training	Complex compared to ANN	Simpler one

8. RESULTS

Performance comparisons of proposed methodologies SVM, Fuzzy SVM, ANN, Fuzzy ANN and existing methodology SLA based mode

Response Time Comparison
Average response time is expressed as the average time taken to respond to the cloud user request from the time of submission to the beginning time its task execution. The response time of the proposed research methodology should be less for the better system performance than the existing research methodology. The comparison chart of the response time of the proposed research methodology with the existing work is depicted in Figure 1.
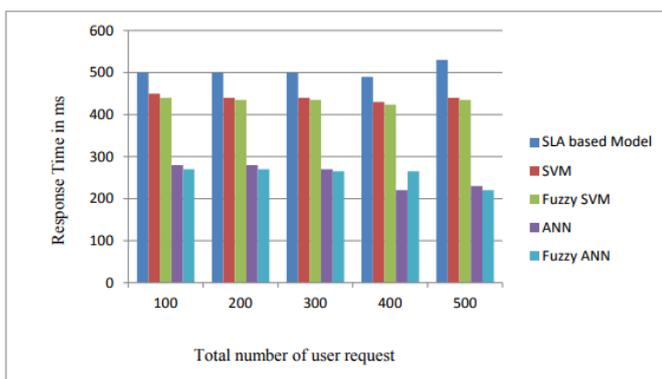

Figure 1. Performance comparisons of proposed methodologies SVM, Fuzzy SVM, ANN, Fuzzy ANN and existing methodology SLA based mode.

Figure 1. Demonstrates that the ANN accomplishes at all reaction time and tolerating more number of clients with less number of VMs. At the point when the entry rate is 100, the contrast between reaction time from ANN and SLA demonstrate based is 44% and that of ANN and SVM is 37%. SLA based model and SVM have comparative reaction times. However, there is a drastic increase in response time when the arrival rate is 500 because more requests are accepted per VM which delays the processing of requests (ANN is 56% less than SLA based model while 47% less than SVM). In case of average response time, fuzzy SVM attain slightly better response time when compared to SVM. And the same scenario is found in case of fuzzy ANN.

9. CONCLUSION

Cloud resource provisioning is a highly sensitive task which needs to be done with greater concern for increasing the user satisfaction level as well as the profit level of the cloud service providers. Admission control is one of the most popular technologies which can perform optimized resource provisioning in the presence of heterogeneous environment with varying user requests. Better and effective admission control task can lead to improved utilization of cloud resources with increased profit for cloud service providers. This is achieved by introducing various research methodologies that attempt to allocate the resources with the concern for increased profit and reduced makespan. The overall findings of this research work conclude that the proposed research methodologies can accomplish the admission control process in the cloud computing environment to the satisfaction of user-specified constraints. The experimental tests conducted to prove that the proposed research namely DLAPUJS provides better results than the existing and previous research works regarding increased profit and decreased the number of VMs initialized as well as execution time with reduced data transfer cost.

10. REFERENCES

[1] Ye Hu, Johnny Wong, Gabriel Iszlai & Marin Litoiu, "Resource Provisioning for Cloud Computing", IBM Canada Ltd. Pp. 101-111 November .2009.

[2] Bhavani, BH & Guruprasad, HS,"Resource Provisioning Techniques in Cloud Computing Environment: A Survey". International Journal of Research in Computer and Communication Technology, vol. 3, no 3. Pp. 395-401, April. 2014.

[3] Mardente, G, Mellia, M & Casetti, C, "Admission Control and Path Allocation for SLAs in DiffServ Networks", High Performance Switching and Routing, 2004. HPSR. Workshop on. April .2004.

[4] Rahul Garg, Ramandeep Singh Randhawa, Huzur Saran & Manpreet Singh,"A SLA Framework for QoS Provisioning and Dynamic Capacity Allocation", Quality of Service, 2002. Tenth IEEE International Workshop on. August.2002.

[5]. Vipul Mathur, Preetam Patil, Varsha Apte & Kannan M Moudgalya,"Adaptive admission control for web applications with variable capacity, In Quality of Service, 2009. IWQoS. 17th International Workshop on, pp. 1-5. IEEE.

[6]   Jaideep, D, Maheshwari, N & Varma, V, "Learning based opportunistic admission control algorithm for MapReduce as a service.Proceedings of the 3rd India Software Engineering Conference", Feb.25-27, ACM Press, New York, USA., pp. 153-160. DOI:10.1145/1730874.1730903

[7]   Gaurav Pandey, Himanshu Gupta & Pabitra Mitra, "Stochastic Scheduling of Active Support Vector Learning Algorithms",Proceedings of the 2005 ACM symposium on Appliedcomputing, pp. 38-42. 2005.

[8]   Josep L Berral, ÍñigoGoiri, Ramón Nou, FerranJulià, JordiGuitart,RicardGavaldà & Jordi Torres, "Towards energy-aware scheduling in data centers using machine learning", Proceedings of the 1st International Conference on Energy-Efficient Computing and Networking, pp. 215-224. April. 2010.

[9]   Sivadon Chaisiri, Bu-Sung Lee & Dusit Niyato, „Optimization of Resource Provisioning Cost in Cloud Computing", IEEE Transactions On Services Computing, vol. 5, no. 2. Pp. 164-177, Feb.2012.

[10]   Marjan Gusev & Sasko Ristov, "The optimal resource allocation among virtual machines in cloud computing", In Proc. of 3rd Int. Conf.on Cloud Computing, GRIDs, and Virtualization (CLOUD COMPUTING 2012), pp. 36-42. 2012.

[11]   Giridas, KL & Shajin Nargunam, A, "Optimal resource allocation technique (ORAT) for green cloud computing", International Journal of Computer Applications vol. 55, no. 5. Pp 20-26. October 2012.

[12]   Bo An, Chunyan Miao & Zhiqi Shen, "Market Based Resource Allocation with Incomplete Information Proceedings of the 20th International Joint Conference on Artificial Intelligence, Hyderabad, India. Pp.1193-1198. Jan.2007.

[13]   Kevin Power, Sunil Vadgama & Chris Williams, "Group Resource Allocation Mechanism for Real-time Services", IEEE C802.16m-08/786.

[14]   ThirumalaRao, B & Reddy, LSS, „Scheduling Data Intensive Workloads through Virtualization on MapReduce based Clouds", Global Journal Computer Science & Technology vol. 11, no. 8, pp. 81-87. August .2011.

[15].   Linlin Wu, Saurabh Kumar Garg & RajkumarBuyya 2012, „SLA-based admission control for a Software-as-a-Service provider in Cloud computing environments"Journal of Computer and System Sciences vol. 78, no.5, pp. 1280-1299. September. 2012.