EXPERIMENTAL STUDY ON CLOUD RESOURCE PREDICTION AND ALLOCATION USING BAT ALGORITHM

PRASADU PEDDI
Research Scholar, Sathyabama University, Chennai, India.

Dr. SIVABALAN ARUMUGAM
Professor, Dept of CSE, Sathyabama University, Chennai, India.

ABSTRACT:

The primary objective of BAT algorithm for resource prediction and allocation in cloud computing environment achieves a better allocation of resources, based on its processing capability, electric power and network bandwidth. BAT enhances the resource scheduling process in cloud computing environment by employing multitasking based resource scheduler. BAT Technique improves the resource scheduling with the adoption of a sub-optimization process to resolve the cloud computing problem. BAT achieves both the allocation of resources as well as the prediction of the system resource; this paper shows experimental analysis of BAT algorithm and Machine learning ANN and SVM.

KEYWORDS: *CSP, QOS, resource allocation, ANN, SVM.*

INTRODUCTION:

Generally, cloud computing is defined as the maintaining and delivering of the most useful services such as software, application, and storage services to users as per their requirements in the best possible useful modes. It reduces the burden of users from keeping the data in their host and maintaining them periodically. Cloud service providers (CSP) are the ones who set up the cloud servers and allocate the required services to the users.

Resource allocation is the process of selecting and providing the most suitable resources that satisfy user requirements. Resource allocation is the most complicated task in the cloud computing environment where the selection of supplies with the satisfaction of Quality of Service (QoS) would be more challenging of the functions. The resources need to be utilized well in achieving higher profit for the cloud services providers. A lot of research work has conducted on optimal resource allocation [9].

The users can utilize the resources that are allocated by the cloud computing service providers through virtualization environment [12]. Virtualization enables service providers to handle multiple users‟ requests simultaneously. Load level of virtual machines needs to be balanced for better resource utilization so that the tasks can completed in a competent manner.

LITERATURE REVIEW:

[1] Proposed SLA based framework for QoS allocation and dynamic capacity allocation. This methodology works based on two types of pricing strategies. These are long-term pricing and short-term pricing. 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 runtime. The overall research of this work provides a continuous allocation of resources to the cloud user with guaranteed SLA.

[2] Designed a novel entity called a domain controller involving cloud tasks and cloud service providers to perform the admission control process. Domain controller will perform admission control tasks by analysing 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.

[3] 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 various 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 assigned to the shared resource that satisfies QoS constraints. After this initial allocation, the data flow rate will calculated regarding QoS constraints based on which resource allocation would be redone to achieve it. This group based resource allocation strategy produces promised optimal resource allocation scenario with the improved satisfaction of user QoS constraints.

[4] solved the problem of grid sites scheduling process with the comfort of QoS parameter values. This optimal grid resource allocation process achieves better utilization by allocating the tasks to resources that can process well with the promising feature of scalability. This approach works well by making use of reserved computing resources and the light path scheme which is utilized well by introducing the joint scheduling approach.

PROBLEM STATEMENT

The various research issues found in previous research methodologies 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 the best fitness value in the optimization approaches. Number of regulations will lead to more computation overheads.
- Data transfer cost would be more in the case of the arrival of the large volume of user requests with different input data requirement that located in different places. The prioritized users are not given importance in previous methodologies.

METHODOLOGY:

Machine learning based approach for resource prediction and allocation with BAT algorithm, this research considers two machine learning approaches for resource allocation, namely Support vector machine (SVM) classification approach [7] and the Artificial Neural Network (ANN) classification approach. In this approach, QoS constraints like profit and the makespan considered for improved user satisfaction level.

Position Balanced Parallel Particle Swarm Optimization (PB-PPSO) [2] method utilized for optimising the resource allocation process. This approach extracts rules from the training set and then finds the optimal control based on the best fitness values. By using these rules, optimal resource allocation with the satisfaction of profit and makespan parameters ensured for the current distribution.

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 functions and resources together. After clustering, optimal resource allocation for those clusters done by using the Position Balanced Parallel Particle Swarm Optimization approach.

Data Locality-Aware Prioritized User Job Scheduling (DLAPUJS). This approach can efficiently handle the large volume of user-submitted tasks and allocate the data required for completing the work. It will categorize the user's requirements in terms of their resource requirement by using the K-Nearest Neighbour (KNN) classification algorithm. After classification, users would be prioritized depending on their resource requirements, and the resource allocation would be done for the users' tasks based on their priority level. Finally, the data required for execution would locate and retrieved from the path in which data transfer cost can reduced considerably.

EXPERIMENTAL RESULT

In this research, optimal resource provisioning for the set of tasks is done by introducing different approaches. These resource allocation techniques were implemented using the CloudSim, a Cloud environment simulator. A total of 500 user requests e considered for the research with a minimum of 100 virtual machines available for performance evaluation. The performance of the proposed methodology DLAPUJS method compared to previous methods like NMF-PB-PPSO, PB-PPSO, ANN and SVM from the user and the resource provider perspectives. On the user side, the number of user requests accepted considered. In the experimental results, three performance metrics such as total profit in $, Average Response Time in milliseconds and a total number of initiated VM‟s compared. The actual values that obtained while evaluating the proposed research methodologies are listed down in **Table 1.** Performance comparison of proposed methodologies "SVM, ANN, PB-PPSO, NMF-PB-PPSO, DLAPUJS."

Number of user requests	Profit in $					Average Response Time in ms				
	SVM	ANN	PB-PPSO	NMF-PB-PPSO	DLAPUJS	SVM	ANN	PB-PPSO	NMF-PB-PPSO	DLAPUJS
100	890	980	1510	1800	2450	450	280	215	165	115
200	2050	2500	2680	2910	3200	440	280	225	180	139
300	3800	4000	4300	4560	5800	440	270	260	195	155
400	4900	5700	5850	6100	8900	430	220	190	170	162
500	5900	6800	7100	7400	10100	440	230	205	198	168

Number of user requests	Number of VM initiated					Data Transfer Cost in $				
	SVM	ANN	PB-PPSO	NMF-PB-PPSO	DLAPUJS	SVM	ANN	PB-PPSO	NMF-PB-PPSO	DLAPUJS
100	15	10	7	7	7	370	290	265	250	190
200	20	15	12	12	11	510	490	420	380	330
300	25	15	11	11	11	470	440	440	430	375
400	30	20	13	13	10	550	550	530	480	420
500	40	25	19	20	17	680	620	590	560	445

Table 1: Performance comparison of proposed methodologies "SVM, ANN, PB-PPSO, NMF-PB-PPSO, DLAPUJS."

PROFIT COMPARISON

Profit defined as the difference between the total revenue of return and the total processing cost.

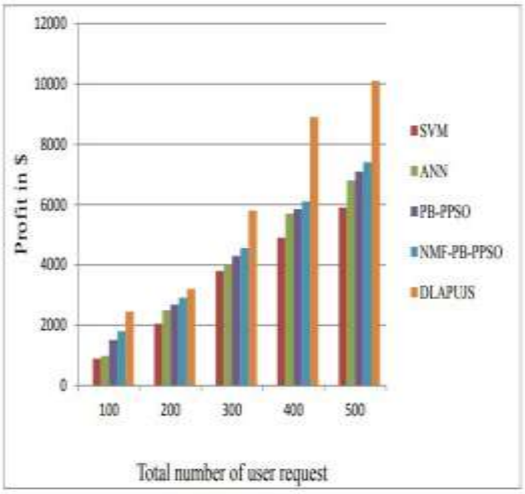

Fig1: Comparison chart of profit

Figure 1 shows that the DLAPUJS method achieves higher advantage and initiating the least number of VMs when the arrival rate increased from 100 to 500. When the user request number is increased, the total profit is risen by the proposed method when compared to the existing system. This is because when the number of requests is increased, the number of users being accepted is increased too by utilizing initiated VMs. From this graph, the final research methodology seeks to provide the better results than the previous methodologies. When the number of user request is 500, DLAPUJS shows improved results of 71 % better than SVM, 48 % better than ANN, 42 % better than PB-PPSO, 36 % better than NMF-PB-PPSO.

RESPONSE TIME COMPARISON

Average response time is defined as the average time taken for responding to the customer request for cloud service.

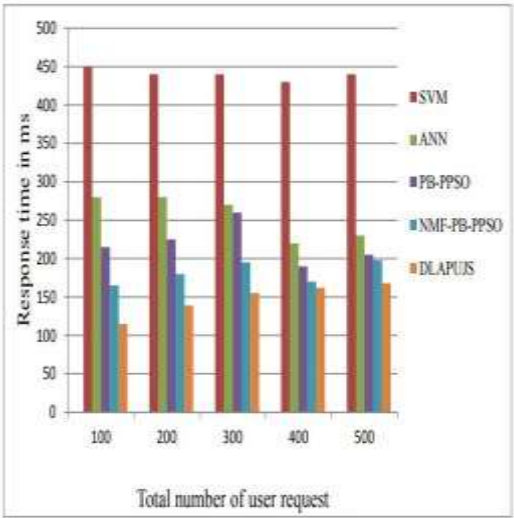

Figure 2 shows that the DLAPUJS method achieves less average response time and initiates the least number of VMs when the arrival rate is increased from 100 to 500. When the user request number is increased, the average response time is decreased in proposed method as compared to the existing system. From this graph, final research methodology concludes to provide better result than the previous methodologies and the DLAPUJS shows improved results of 74% better than SVM, 59% better than ANN, 46 % better than PB-PPSO, 30% better than NMF-PB-PPSO when the number of user request is 100.

CONCLUSION:

The proposed work achieve better results with metrics, like interoperability, computational cost, scalability, load balance factor, clustering efficiency, generic based cloud services, multi-task clustering effect, communication cost and computational complexity. More specifically, IARA Technique reduces the interoperability rate by about 11-27% in interference minimization when compared with the existing methods. Similarly, ALB Approach provides better load balance factor of about 8 – 25 % in contrast to the other existing schemes. Further, GCWM Scheduler improves the multitasking Clustering Effect by 17 – 25 % when compared with other existing works.

REFERENCES

1. *Ajith Singh and hemalatha, "An Approach on Semi-Distributed Load Balancing Algorithm for Cloud Computing System", International Journal of Computer Applications (0975 – 8887) Volume 56– No.12, October 2012.*
2. *Beaumont, Olivier ; Eyraud-Dubois, Lionel ; Thraves Caro, Christopher ; Rejeb, Hejer "Heterogeneous Resource Allocation under Degree Constraints" Parallel and Distributed Systems, IEEE Transactions on (Volume:24 , Issue: 5) May 2013 Page(s): 926 – 937.*
3. *En-Hao Chang., Chen-Chieh Wang., Chien-Te Liu., Kuan-Chung Chen., and Chung-Ho Chen, "Virtualization Technology for TCP/IP Offload Engine," Journal Of Latex Class Files, Vol. 11, No. 4, December 2013.*
4. *Harpreet Kaur and Maninder Singh, "A Task scheduling and Resource Allocation Algorithm for Cloud using Live Migration and Priorities", International Journal of Computer Applications (0975 – 8887), Volume 84 – No 13, December 2013.*
5. *Imad M. Abbadi., and Anbang Ruan., "Towards Trustworthy Resource Scheduling in Clouds," IEEE Transactions on Information Forensics and Security, Vol. 8, No. 6, June 2013.*
6. *Junwei Cao, Keqin Li, Ivan Stojmenovic, "Optimal Power Allocation and Load Distribution for Multiple Heterogeneous Multicore Server Processors across Clouds and Data Centers", IEEE Transactions on Computers, Vol. 63, No. 1, January 2014.*
7. *Jachak K.B., Korde S.K., Ghorpade P.P. And Gagare G.J., "Homomorphic Authentication with Random Masking Technique Ensuring Privacy & Security in Cloud Computing, "BIOINFO Security Informatics ISSN: 2249-9423 & E-ISSN: 2249-9431, Volume 2, Issue 2, 2012, pp.-49-52.*
8. *Konstantinos Tsakalozos., Mema Roussopoulos., and Alex Delis., "Hint-Based Execution of Workloads in Clouds with Nefeli," IEEE Transactions on Parallel and Distributed Systems, Vol. 24, No. 7, July 2013.*
9. *Nallur Bahsoon, "A Decentralized Self-Adaptation Mechanism for Service-Based Applications in the Cloud" Software Engineering, IEEE Transactions on (Volume: 39, Issue: 5) 591 – 612 July 2012.*
10. *Siddha S., Pallipadi V., Mallick A., "Process Scheduling Challenges in Multicore Processors", Intel Technology Journal, Vol. 11 , Nov 2011, Pages 15-18, 2007.*
11. *Yan Zhu., and Shanbiao Wang., "Secure Collaborative Integrity Verification for Hybrid Cloud Environments," International Journal of Cooperative Information Systems Vol. 21, No. 3 165–197. DOI: 10.1142/S0218843012410018, (2012).*
12. *Zhen Xiao; Weijia Song; Qi Chen "Dynamic Resource Allocation Using Virtual Machines for Cloud Computing Environment" Parallel and Distributed Systems, IEEE Transactions on (Volume: 24, Issue: 6) June 2013 Page(s): 1107 – 1117*