Optimal Scheduling of Vms in Queuing Cloud Computing Systems With A Heterogeneous Workload

Allen S.V[1], Sudhakar B[2], Dr. G Ayyappan[3], Reena R[4]
[1, 2] Dept of Computer Science and Engineering
[3, 4]Professor, Dept of Computer Science and Engineering
[1, 2, 3, 4] Prince Shri Venkateshwara Padmavathy Engineering College, Chennai, Tamil Nadu

Abstract- *The applications can be run remotely using other people's servers. This is done with a simple user interface or an application format. In general, Cloud is simply the Internet which in turn is a network of remote servers. 'Pay only for what you use' is the bottom line for computing in cloud environment. The Cloud is a pool of heterogeneous resources. The storage dealt with cloud computing has opened the gates to Load Balancing. Load Balancing refers to the process of balancing the load efficiently among the servers such that no server is overloaded or under loaded. The prime motive of this project is to minimize the response time thereby increasing the performance. This is implemented by using the Improved Weighted Round Robin dividing the workload efficiently and successfully, thereby multiplying the scalability and the performance of computing, subtracting the response time. Ant Colony Optimization is basic foraging behaviour of an ant that encouraged them to find the optimal shortest path and quality of services.*

Keywords- Load balancing, Round Robin, Ant colony optimization, Optimal shortest path, Quality of services

I. INTRODUCTION

The prime motive of this project is to minimize the response time thereby increasing the performance. This is implemented by using the Improved Weighted Round Robin dividing the workload efficiently and successfully, thereby multiplying the scalability and the performance of computing, subtracting the response time. Ant Colony Optimization is basic foraging behaviour of an ant that encouraged them to find the optimal shortest path and quality of services. This paper focused on the key techniques of application software service in campus science cloud computing environment, in this situation, the software licenses and hardware are centralized management. All the hardware and software resources are shared based on user roles and priority. The algorithm of software and hardware co-allocation is presented. The evaluation results prove that it can improve resource utilization, meet the role-based service level agreements, maintain system load balancing, etc. We use three types of algorithms such as Genetic algorithm, Analytic algorithm, Ant optimization algorithm.

II. LITERATURE SURVEY

In C. Wilson, H. Ballani, T. Karagiannis, and A. Rowtron, "Better never than late: Meeting deadlines in datacenter networks," (2011) **[1]**The soft real-time nature of large-scale web applications in today's datacenters, combined with their distributed workflow, leads to deadlines being associated with the datacenter application traffic. A network flow is useful, and contributes to application throughput and operator revenue if, and only if, it completes within its deadline. Today's transport protocols (TCP included), given their Internet origins, are agnostic to such flow deadlines. Instead, they strive to share network resources fairly. We show that this can hurt application performance. Motivated by these observations, and other (previously known) deficiencies of TCP in the datacenter environment, this paper presents the design and implementation of D3, a deadline-aware control protocol that is customized for the datacenter environment. D3 uses explicit rate control to apportion bandwidth according to flow deadlines. Evaluation from a 19-node, two-tier datacenter testbed shows that D 3 , even without any deadline information, easily outperforms TCP in terms of short flow latency and burst tolerance. Further, by utilizing deadline information, D3 effectively doubles the peak load that the datacenter network can support.

In A.D.Papaioannou, R. Nejabati, and D. Simeonidou, "The benefits of a disaggregated data centre: A resource allocation approach," (2016)**[2]**Disaggregation of IT resources has been proposed as an alternative configuration for data centres. Comparing to the monolithic server approach that data centres are being built now, in a disaggregated data centre, CPU, memory and storage are separate resource blades and they are interconnected via a network fabric. That brings greater flexibility and improvements to the future data centers in terms of utilization efficiency and energy consumption. The key enabler for the disaggregated data center is the network, which should support the bandwidth and latency requirements of the communication that is currently inside the server. In

addition, a management software is required to create the logical connection of the resources needed by an application. In this paper, we propose a disaggregated data center network architecture, we present the first scheduling algorithm specifically designed for disaggregated computing and we demonstrate the benefits that disaggregation will bring to operators.

III. EXISTING SYSTEM

The existing systems have worked on making complete utilization of resources and improving the response time of the job, looking into the load balancing process that holds the condition where some of the nodes involved are overloaded while others are not. Therefore, load balancing algorithms are chosen appropriately that allows selection of virtual machines or servers. All of the above-mentioned factors collaboratively make up an efficient load balanced cloud environment with suitable algorithms that support the features.

IV. PROPOSED SYSTEM

We mainly discuss three algorithm we developed a new generalized priority-based algorithm with limited task, future we will take more task and try to reduce the execution time as presented and we develop this algorithm to grid environment and will observe the difference of time in cloud a grid. We propose a heterogeneous resource allocation approach, called skewness-avoidance multi-resource allocation (SAMR), to allocate resource according to diversified requirements on different types of resources. Our solution includes a VM allocation algorithm to ensure heterogeneous workloads are allocated appropriately to avoid skewed resource utilization in PMs, and a model-based approach to estimate the appropriate number of active PMs to operate SAMR.

V. DESIGN AND ANALYSIS

The optimal scheduling of VMs in cloud computing system is a combination of various steps, techniques, and algorithms which are combined in a controlled manner. The proposed system has various modules through which the scheduling can be done.

SERVICE AND VM SCHEDULING

A scheduling framework can be implemented by using different parameters. Good scheduling framework should include the following specifications. It must focus on: Load balancing and energy efficiency of the data centers and virtual machines. Quality of service parameters calculated by the user which contain execution time, cost and so on. It should satisfy the security features. Fairness resource allocation places a vital role in scheduling.

ANALYSIS

In this module, we mainly discuss SAMR algorithm we developed a new generalized priority based algorithm with limited task, future we will take more task and try to reduce the execution time as presented and we develop this algorithm to grid environment and will observe the difference of time in cloud an grid.

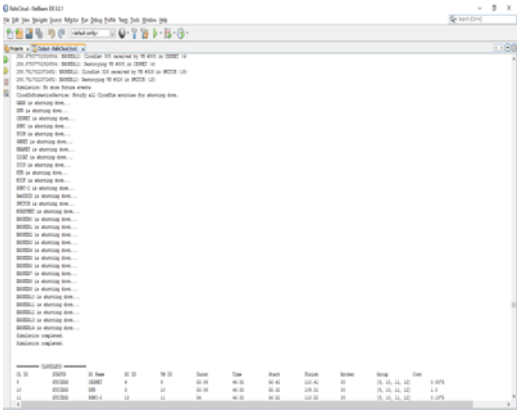

Figure 1: CLOUDSIM ANALYSIS

STATIC LOAD BALANCING

In static algorithm the processes are assigned to the processors at the compile time according to the performance of the nodes. Once the processes are assigned, no change or reassignment is possible at the run time. Number of jobs in each node is fixed in static load balancing algorithm. Static algorithms do not collect any information about the nodes . The assignment of jobs is done to the processing nodes on the basis of the following factors: incoming time, extent of resource needed, mean execution time and inter-process communications. Since these factors should be measured before the assignment, this is why static load balance is also called probabilistic algorithm.

OPTIMAL STATIC LOAD BALANCING ALGORITHMS

If all the information and resources related to a system are known optimal static load balancing can be done. It is possible to increase throughput of a system and to maximize the use of the resources by optimal load balancing algorithm.

QUALITY OF A LOAD BALANCING ALGORITHM

Quality of a load balancing algorithm is dependent on two factors. Firstly, number of steps that are needed to get a balanced state. Secondly the extent of load that moves over the link to which nodes are connected.

VI. RESULT

In this module, result indicate that our model increase utilization of global scheduler and decrease waiting time. And also indicated that model decrease waiting time at global scheduler in cloud architecture.

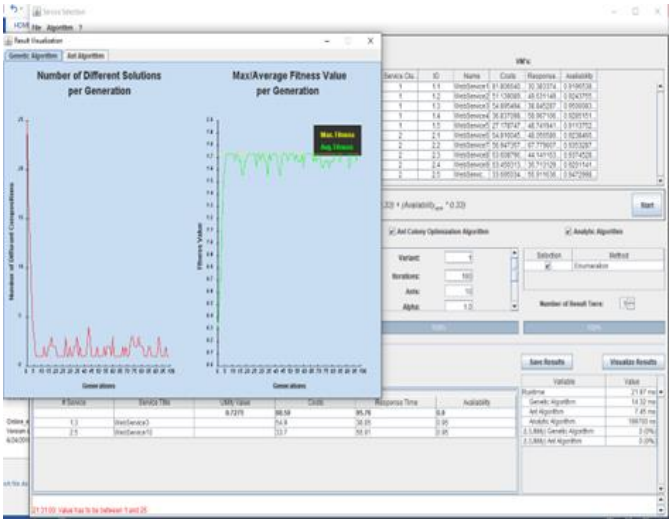

Figure 2: RESULT

VII. CONCLUSION

Design of a resource management system for cloud computing services, implementation and it presented and evaluated. Based on the changing demands of adaptively multiplexing physical resources, a system of virtual us. As appropriate to the capacity of the server is fully utilized, this are using a Utility function that combines the VM resources and different characteristics. The algorithm has been achieved both of green computing for a system with multi-resource constraints and avoid overload.Future work may includes improving our algorithm in the specific data center network topologies with energy consumption of switches considered.

REFERENCES

[1] C. Wilson, H. Ballani, T. Karagiannis, and A. Rowtron, "Better never than late: Meeting deadlines in datacenter networks," SIGCOMM Compute. Commune. Rev., vol. 41, no. 4, pp. 50–61, 2011.

[2] A. D. Papaioannou, R. Nejabati, and D. Simeonidou, "The benefits of a disaggregated data center: A resource allocation approach," in Proc. IEEE GLOBECOM, pp. 1–7, Dec 2016.

[3] A. Tchernykh, U. Schwiegelsohn, V. Alexandrov, and E. ghazaliTalbi, "Towards understanding uncertainty in cloud computing resource provisioning," in Proc. ICCS, pp. 1772–1781, 2015.

[4] J. Hu, J. Gu, G. Sun, and T. Zhao, "A scheduling strategy on load balancing of virtual machine resources in cloud computing environment," in Proc. PAAP, pp. 89–96, 2010.

[5] K.-M. Cho, P.-W. Tsai, C.-W. Tsai, and C.-S. Yang, "A hybrid metaheuristic algorithm for vm scheduling with load balancing in cloud computing," Neural Compute. Appl., vol. 26, no. 6, pp. 1297– 1309, 2015.

[6] S. Rampersaud and D. Grosu, "Sharing-aware online virtual machine packing in heterogeneous resource clouds," IEEE Transactions on Parallel and Distributed Systems, vol. 28, pp. 2046–2059, July 2017.

[7] S. S. Rajput and V. S. Kushwah, "A genetic based improved load balanced min-min task scheduling algorithm for load balancing in cloud computing," in 2016 8th International Conference on Computational Intelligence and Communication Networks (CICN), pp. 677–681, 2016.

[8] S. T. Maguluri, R. Srikant, and L. Ying, "Stochastic models of load balancing and scheduling in cloud computing clusters," in Proc. IEEE INFOCOM, pp. 702–710, 2012.

[9] S. H. H. Madni, M. S. A. Latiff, Y. Coulibaly, and S. M. Abdulhamid, "Resource scheduling for infrastructure as a service (iaas) in cloud computing: Challenges and opportunities," Journal of Network and Computer Applications, vol. 68, no. Supplement C, pp. 173–200, 2016.

[10] J. Ma, W. Li, T. Fu, L. Yan, and G. Hu, A Novel Dynamic Task Scheduling Algorithm Based on Improved Genetic Algorithm in Cloud Computing, pp. 829–835. New Delhi: Springer India, 2016.