Enhanced Weighted Round Robin Load Balancing Algorithm in Cloud Computing

Shivangi Mayur, Nidhi Chaudhary

Abstract: The cloud/utility computing model requires a dynamic task assignment to cloud sites with the goal that the performance and demand handling is done as effectively as would be prudent. Efficient load balancing and proper allocation of resources are vital systems to improve the execution of different services and make legitimate usage of existing assets in the cloud computing atmosphere. Consequently, the cloud-based infrastructure has numerous kinds of load concerns such as CPU load, server load, memory drain, network load, etc. Thus, an appropriate load balancing system helps in realizing failures, reducing backlog problems, adaptability, proper resource distribution, expanding dependability and client fulfillment and so forth in distributed environment. This thesis reviewed various popular load balancing algorithms. Modified round robin algorithms are popularly employed by various giant companies for scheduling issues and load balancing. An enhanced weighted round robin algorithm is discussed in this paper concentrating on efficient load balancing and effective task scheduling and resource management.

Keywords: Cloud computing, Load balancing algorithms, Resource allocation, Task scheduling.

I. INTRODUCTION

A. Background

Cloud computing today is a rapidly expanding field in the computing and software industry. Three significant facilities offered by the cloud are IaaS, SaaS, and PaaS. As the Cloud develops, new opportunities are opened up on how programs can be made on the internet and how various facilities can be supplied to end users via virtualization. The cloud computing environment needs two different methods for traditional service providers, facilities and service providers. Cloud platforms agreement of infrastructure providers and uses resources accordingly. Service providers give end-users support services from infrastructure providers. The business giants, such as Google, Microsoft and Amazon are seen as a major impact of the next generation of technology in today's information technology companies attracted towards Cloud computing.

The elastic, scalable and simple way to maintain and collect information and documents as part of its facilities is offered in Cloud Computing. Especially for the availability of large datasets and files to a growing amount of customers worldwide. Many methods for optimizing and optimizing operations and providing users with satisfactory performance levels are needed for dealing with these kind of leading datasets. Dynamic load balance or task management are one of the main problems correlated with cloud computing. Cloud computing requires mainly that work cycles be assigned to cloud servers so as to monitor the requests and demands as effectively as feasible (Al Nuaimi et al. 2012), while at the same time being willing to meet the multiple limitations of stratification and elevated communication costs and delays.

A number of load balancing algorithms in the cloud computing are existing for the assignment of the effective designation for virtual machines (VMs). Among such existed algorithm which is to be utilized is the main decision is to be considered. Some of those algorithms have been explained in this paper. Hence for having accurate utilization of resources and being faithful with all the resources, idea of load balancing is being conducted.

B. An Overview of Cloud Computing

A cloud computing can be recalled as an internet-based framework in which the computing resources such as storage, hardware, databases, software applications, networks and even integral operating systems in conception of information technology (IT) services are available to the end clients as per-the-demand of services. Moreover, there not much inclusion of new technologies in cloud computing; yet, it empowers to increase the determinative leads in managing IT services and liberate cost issues (Baroncelli, Martini, and Castoldi 2010)(Lehrig, Eikerling, and Becker 2015).

### C. Cloud Computing and Load Balancing Algorithms

In any system, one of the important factor is the load and hence the load balancing becomes significant which manages the Load across all system. In this system it is the load that gets fragmented and the resources gets distributed in such a way that load is shared equally among all over the system. This results into faster work and lesser hindrance in the work. But the question is that how to get to such system and the answer is the availability of methods to balance requests and to offer the solution of fast reply for request. Here what happens is that online traffic is maintained through the workload dissemination. It is made to happen on

its own among the different servers. There is an increment in the throughput whereas decrement in the time for response and ultimately results into overload reduction.
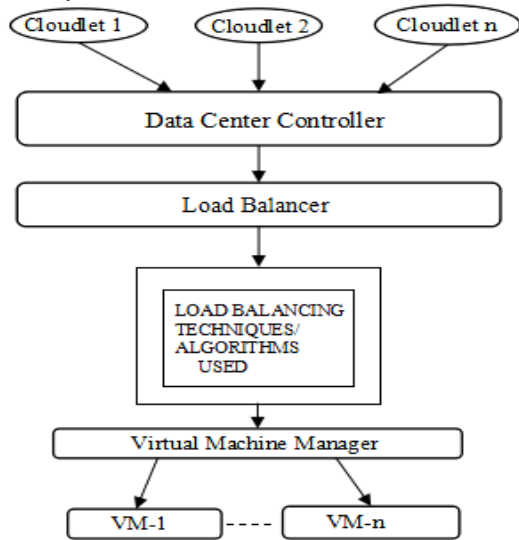
Fig. 1.1: Schematic representation of execution of a load balancing algorithm.

D. Categories of Load Balancing Algorithms

Load balancing algorithms are broadly divided into two major categories (Ghomi, Rahmani, and Qader 2017):

• Based on how the load is distributed and how processes are allocated to system nodes.

• Based on the information status of the nodes.

E. Existing Load Balancing Algorithms

The monitoring solutions to the Load balancing can be split into load balancers based on software and load balancers based on hardware. Load balancers based on the hardware equipments have the capacity to manage high-speed traffics on the network. These are specialized containers, including application particular integrated circuits (ASICs) tailored to a particular application whereas conventional hardware parts and normal working systems are used by load-balancers based on software. Distributed network connections to obtain minimum response time and peak performance and prevent overload is the key feature of cloud computing. For the load distribution, there are various load balance algorithms.

II. LITERATURE REVIEW

An extensive study on various literatures has been performed to conceptualize the process of load balancing under the cloud environment. The research exploration is grouped under various categories based on the nature of the implemented system design.

A. Cloud Computing and Load Balancing

According to (Rahman et al. 2014) Load Balancer as a Service (LBaaS) is one of the more popular methods; rather than building a load balancer for each cloud, companies prefer to consume LbaaS. In this method, the load balancer will be utilized through the network, and it requires information about the entire network, including information about every node within it. The majority of LbaaS providers uses round-robin or weighted round-robin algorithm as their static algorithm.

Priyanka Singh, (2016), in her detailed assessment of the different algorithms suggested previously in the paper "Assorted Load Balancing Algorithms in Cloud Computing: A Survey." An investigation in the multiple algorithms has been performed to discover an appropriate load balancer alternative in the cloud computing ecosystem. In the survey paper the author discussed the advantages and disadvantages of various load balancing algorithms.

In order to maximize output and to balance the objectives of VM workloads, a load balancing algorithm based on Honey Bee Behavior was proposed (Krishna, 2013) with the purpose of achieve a well-balanced load over VM. The time to wait for the jobs for their turn in the queue is therefore small. The use of this algorithm has enhanced the median run time and the average waiting time of queue assignments. This algorithm serves to compensate for non-preventative, unbiased tasks and heterogeneous technologies.

In another research by (Zalavadiya and Vaghela, 2016), an algorithm was proposed to achieve minimum completion time by considering priority as a factor. In this proposed honey bee algorithm, when the task scheduler finds unevenness in the load among the nodes in the network, it migrates the high priority task from the overloaded VM to any VMs with a lower number of similar priority tasks. This algorithm also uses the preemptive method of migrating dependent tasks. This results in improving the minimum completion time and maximizing utilization.

A novel algorithm is proposed by (Gursharan et al. 2015) for memory reusing during VM migrations. A VM migration helps in balancing the load and maximizes the resource utilization. It will also lead to high memory usage in storing the background data before the transfer. This algorithm instead of transferring the data stores it in the original host and reuses the same once when the VM is migrated back to the original host. Results have proven that memory consumption is reduced by 33% when the size of the memory image is reduced and when restored in the original host.

III. METHODOLOGY

A. Background

Load balancing means a mechanism in which a single machine divides up the amount of work into two or even more computer systems that allow the user to perform their work faster and each computer involved completes its job in a similar amount of time. In terms of software, hardware, or with a mixture of the two, load balancing can be attained. Throughout the load balancing phase, several servers are engaged. This is usually performed by scripting them using different programming algorithms to efficiently assign workloads. The working load is allocated and balanced by means of Load Balancers between two or multiple cloud

servers. The technology can therefore be designed to meet bumps in the activity, automating resource allocations to ensure minimum delays and time responses. In instances requiring assured service continuity, a heavy traffic processing and experiencing unexpected demand peaks, a load balancer is strongly recommended. The amount of functioning cloud servers can prevent overloads.

B. Cloud Scheduler and Load Balancer

The scheduler and load balancer are the important components of the network that use a load balancing algorithm to distribute the loads evenly across the resources. The scheduler is a component that helps in deciding the appropriate VM to which an arrived task should be allocated. It ensures that the task is assigned to a VM that takes less time to complete the task by considering the amount of load in that specific VM and the total time needed to complete the entire load (Hu, 2010). The scheduler makes use of the information about the resources from the resource manager to find the appropriate VM to handle the incoming task. Each VM has an in-coming task queue, and the tasks allocated to each VM is available in the incoming task queue.

The load balancer is another important component in this system; it ensures that the load in the cloud network is evenly distributed. If a load balancer identifies any of the VMs as overloaded, it immediately migrates the task from an overloaded VM to an under-loaded or ideally loaded VM. An incoming task first reaches the task manager, which identifies the length and priority of the task. The resource manager helps to calculate the weight to each VM by using various information, such as the processing capacity of VMs, the number of tasks in each VM incoming task queue, and total time required to complete the tasks of each priority. The load balancer helps in finding load unevenness in the system by calculating the ratio between the total number of jobs in the queue and the number of available VMs. It decides which VM is less utilized and communicate with the scheduler to allocate the incoming task to that VM. The load on each VM is also measured to ensure it is not overloaded.

C. Algorithms

Round Robin Algorithm

Existing round robin is the algorithm used by the CPU for task scheduling while the process is running. The round robin algorithms are specifically designed for systems for time scheduling and sharing. There are described/defined small time frames also called or quantum or time slices with round robin's circular queue, also recognized as prepared queue, all procedures in this algorithm are maintained. The tail of a waiting / circular queue is applied to every new process. With the use of this algorithm, CPU ensures that the time span (any natural number) is allocated equally and in circular queuing order designated to every process that deals with all processes without any priorities. Thus, it is also called as the cyclical manager. Regardless of the load throughout the VMs, the round robin algorithm assigns the next VM task present on the queue. Round Robin algorithm does not take the resource capacity, priority and duration of assignments into

consideration. The greater priority and the longer assignments therefore leads to greater reaction rates. The primary benefit of the RR algorithm is that it is free of starvation, as compared to FCFS or SRT algorithm. Hence, for a fixed amount of duration (set as time slice), each method shall be performed by the CPU. Therefore, there is no transaction remaining in the queue for the CPU to wait for its execution turn. A round robin algorithm is easily implemented and straightforward algorithm in load balancing applications.

Weighted Round Robin Algorithm

The weighted round robin (WRR) takes into account the resource capacities of the VMs and allocates higher volumes to the VMs, according to the weighting of each VMs. This classifier discovers the suitable VM in the estimation for the time of wait, whenever a fresh entry is obtained. But it subsequently fails to consider the length of the tasks for the selection of the appropriate VM.

Proposed Algorithm

Several algorithms for load balancing, like dynamic load balancer, round robin(RR), weighted round robin(WRR), ESCE algorithm, Throttled algorithm, Ant Colony algorithm , First Come First Serve, are available. Weighted round robin(WRR) and FIFO(First In First Out) form the most commonly utilized planning methods for a non-preemptive scheme. The algorithm proposed in this study is Enhanced weighted round robin(EWRR). The Weighted Round Robin(WRR) algorithm is the basis of this algorithm. The WRR algorithm operates by giving the highest possible requirements for the maximal weight of the server. For instance if two servers are available, one with the server has Intel i7 processor and the other is equipped with Intel i3 processing server and relatively small processing speed, then maximum weight is assigned to the server with best specifications. The WRR algorithm, in brief, is a unique situation of the Round Robin algorithm, working well with various servers.

The EWRR functions in such a way that the server framework is not only taken into account, but also the execution times of the tasks are taken into account in order to assign a task to the server with the maximum time for execution. This guarantees that the distribution of load is uniformly spread between servers, reducing the standard reaction time.

Two variants, V_{max} and V_{min} are used to calculate the threshold range. Threshold for each VM is calculated as follows:

$$T_{min} = \frac{c}{C} * V_{min} * L * n$$

Where V_{min} represents the minimum variance, L represents the total capacity of a node, and n represents the total number of VMs. T_{min} represents the minimum threshold value of a VM.

$$T_{max} = \frac{c}{C} * V_{max} * L * n$$

Here, T_{max} represents the maximum threshold value of a VM and V_{max} represents the maximum variance.

IV. DISCUSSIONS AND RESULTS CONCLUSION

Cloud Computing is a new paradigm with ever increasing technological advancement, emergence tending to rapidly transform the IT landscape. Furthermore, despite the important advantages of the cloud ecosystem, the present model's major problems is its load balancing issues. This article discusses many currently reviewed load balancing methods, which relies primarily on overhead reduction, system performance improvement, improved resource uses customer response time and throughput optimization. In the future, a large numbers of constraints and various kinds of soft-computing methodologies for making better use of resources and handling user demands may be introduced in the cloud system.

Here we studied the efficiency of cloud-based infrastructure based on load balancing factors in various components of cloud computing. The load balancer plays a major role in cloud-based services. The services offered in the cloud communicate directly with the clients within a specific cloud infrastructure. The user-to-cloud communication requires load transition to VMs as well as other infrastructure. Various researchers and cloud designers used swarm based work design and assignment scheduling method to enhance load effectiveness in cloud computing. The non-preventive planning method is like round robin algorithms which are highly effective in sympathizing with the aged and traditional methods like FCFS or round SRT. Moreover, an enhanced weighted round robin algorithm was introduced in this research paper and its performance efficiency was analyzed. It was found that IWRR performed much better in comparison to RR and WRR individually, where migration and time of execution of algorithm was tested in the experiments.

V. FUTURE SCOPE

Future plan is to offer security algorithm with load balancing

REFERENCES

1. Al nuaimi, klaithem, nader mohamed, mariam al nuaimi, and jameela al-jaroodi, 2012. "a survey of load balancing in cloud computing: challenges and algorithms." in 2012 second symposium on network cloud computing and applications, pp. 137-142.
2. Baroncelli, fabio, barbara martini, and piero castoldi, 2010. "network virtualization for cloud computing." *annals of telecommunications-annales des télécommunications,* 65.11-12, pp. 713-721.
3. Ghomi, einollah jafarnejad, amir masoud rahmani, and nooruldeen nasih qader, 2017. "load-balancing algorithms in cloud computing: a survey." journal of network and computer applications 88, pp. 50-71.
4. Gursharan singh, sunny behal, monal taneja, 2015, "advanced memory reusing mechanism for virtual machines in cloud computing", 3rd international conference on recent trends in computing 2015 (icrtc-2015), procedia computer science 57, pp. 91 – 103.
5. Hu j, gu j, sun g, zhao t. 2010, "a scheduling strategy on load balancing of virtual machine resources in cloud computing environment." In 2010 3rd international symposium on parallel architectures, algorithms and programming, pp. 89-96.
6. Krishna pv. 2013, "honey bee behavior inspired load balancing of tasks in cloud computing environments." Applied soft computing. 1;13(5): pp. 2292-303.
7. Lehrig sebastian, hendrik eikerling, and steffen becker. 2015, "scalability, elasticity, and efficiency in cloud computing: a systematic literature review of definitions and metrics." in proceedings of the 11th international acm sigsoft conference on quality of software architectures. Pp. 83-92.
8. Rahman m, iqbal s, gao j. 2014, "load balancer as a service in cloud computing." In 2014 ieee 8th international symposium on service oriented system engineering. Pp. 204-211.
9. Singh priyanka, palak baaga, and saurabh gupta. 2016, "assorted load balancing algorithms in cloud computing: a survey." international journal of computer applications. Vol. 143, no. 7, pp. 34-40.
10. Zalavadiya k, vaghela d, 2016. "honey bee behavior load balancing of tasks in cloud computing." International journal of computer applications. 139(1), pp. 16-9.

AUTHORS PROFILE

Shivangi Mayur pursuing Masters of Engineering in Computer Science from MIET, Meerut, AKTU University, Lucknow, India. One paper published in JETIR Publications. Her researched topic was Load Balancing in Cloud Computing. GATE-2017 qualified.

Nidhi Chaudhary received the Masters of Engineering in Computer Science degree from NITTTR, Chandigarh, in 2017 and She is currently working as an Assistant professor in MIET, Meerut, CSE Department. herPaper published in IEEE, JETIR, IJSCE. GATE -2010 qualified. Certified by IBM, WIPRO & TCS. Gold Medalist in Master's of Engineerings