Energy-Efficient Load Balancing in Cloud: A Survey on Green Cloud

M. Nirmala, Associate Professor, Department of Computer Science & Engineering, Aurora's Technology & Research Institute, Uppal, Hyderabad. Telangana, India.
Dr. T.Adi Lakshmi, Professor, Department of CSE, Vasavi College of Engineering, Ibrahimpatnam, Hyderabad, Telangana, India.

Abstract: With increased use of cloud computing architectures, organizations are trying to reduce the power consumed by unutilized resources. Load Balancing can help in reducing energy consumption by evenly distributing the load and minimizing the resource consumption. The ultimate goal of the service provider is to use the computing resources efficiently and gain maximum profits using an efficient load balancing algorithm. Consumption of resources and conservation of energy are not always a prime focus of discussion in the cloud computing. However, resource consumption can be kept to a minimum with proper load balancing which not only helps in reducing costs but making enterprise greener. Scalability which is one of the very important features of cloud computing is also enabled by load balancing. Hence improving resource utility and the performance of a distributed system in such a way will reduce the energy consumption and carbon footprints to achieve Green computing. Various energy aware load balancing techniques are studied and their corresponding advantages, disadvantages and performance metrics are presented.

Keywords: Cloud computing, Virtualization, Energy-aware Load Balancing.

1. Introduction

Cloud computing is the dynamic provisioning of IT capabilities (hardware, software or services) from third parties over a network.Clouds use virtualization technology in distributed data centers to allocate resources to customers as they need them.Redundancy and disaster recovery capabilities are built into cloud computing environments and on demand resource capacity can be used for better resilience when facing increased service demands. The cloud model is strongly based on the concept of 'Location Independence' and Virtualization. Physical resources can be split into a number of logical slices called virtual machines (VMs). Fundamentally, the provider's computing resources are pooled to serve multiple consumers using a multi-tenant model, with different physical and virtual resources dynamically assigned and reassigned according to the consumer demand.

Load balancing is one of the prerequisites to utilize the full resources of parallel and distributed system. Load Balancing can help in reducing energy consumption by evenly distributing the load and minimizing the resource consumption. It is used to achieve a high user satisfaction and resource utilization ratio, making sure that no single node is overwhelmed, hence improving the overall performance of the system.

In this survey paper there were various load balancing techniques are discussed. The paper is organized as follows. Section 2 describes the motivation for Energy aware solutions. Section 3 describes literature review of various loadbalancing papers and section 4 describes the conclusion.

2.MotivationforEnergyAware Solutions

Proper load balancing can help in utilizing the available resources optimally, thereby minimizing the resource consumption. It also helps in implementing fail-over, enabling scalability, avoiding bottlenecks and over-provisioning, reducing response time etc. All VM load balancing methods are designed to determine which virtual machine is assigned to the next request.

The realization that power consumption of cloud computing centers is significant and is expected to increase substantially in the future motivates the interest of the research community in energy-aware resource management and application placement policies and the mechanisms to enforce these policies.

Load balancing is also required to achieve Green computing in clouds which can be done with the help of the following two factors:

- **Reducing Energy Consumption** – Load balancing helps in avoiding overheating by balancing the workload across all the nodes of a cloud, hence reducing the amount of energy consumed.
- **Reducing Carbon Emission**- Energy consumption and carbon emission go hand in hand. The more the energy consumed, higher is the carbon footprint. As the energy consumption is reduced with the help of Load balancing, so is the carbon emission helping in achieving Green computing.

Consumption of resources and conservation of energy are not always a prime focus of discussion in the cloud computing. However, resource consumption can be kept to a minimum with proper load balancing which not only helps in reducing costs but making enterprise greener. Scalability which is one of the very important features of cloud computing is also enabled by load balancing. Hence improving resource utility and the performance of a distributed system in such a way will reduce the energy consumption and carbon footprints to achieve Green computing. Hence, energy-efficient solutions that can address the high energy consumption, both from the perspective of the cloud provider and the environment are required.

3. Existing Energy-Aware Load Balancing Techniques In Cloud

3.1Energy-aware Load Balancing and Application Scaling for the Cloud Ecosystem: AshkanPaya and Dan C. Marinescu et al.[1] proposed that low average server utilization and its impact on the environment make it imperative to devise new energy-aware policies which identify optimal regimes for the cloud servers and, at the same time, prevent SLA violations. They defined an energy-optimal operation regime and attempted to maximize the number of servers operating in this regime. The Proposed algorithm uses five operating regimes, motivated by the desire to distinguish three types of system behavior in terms of power utilization: optimal, suboptimal and undesirable.A server operating in the optimal regime is unlikely to request a VM migration in the immediate future and to cause an SLA violation and one in a sub-optimal regime is more likely to request a VM migration, while one in the undesirable high regime is very likely to require VM migration. Servers in the undesirable-low regime should be switched to a sleep state as soon as feasible. The objective of the algorithm is to ensure that the largest possible number of active servers operate within the boundaries of their respective optimal operating regime.

Cloud is organized as a set of clusters, where each cluster is managed by a cluster leader. Each cluster leader maintains several data structures like optimal list, watch list, migration list and sleep list. The leader performs functions like admission control, server consolidation and reactivation and VM migration.

The actions implementing this policy are: (a) migrate VMs from a server operating in the undesirable-low regime and then switch the server to a sleep state; (b) switch an idle server to a sleep state and reactivate servers in a sleep state when the cluster load increases; (c) migrate the VMs from an overloaded server, a server operating in the undesirable-high regime with applications predicted to increase their demands for computing in the next reallocation cycles. This algorithm has the advantage that it reduces energy consumption by switching VMs in undesirable low regime to sleep state and overloaded Servers are balanced by VM migration. But, the disadvantage is the maintenance of several data structures and an overhead of updating each list for every task that needs to be executed. Also, VMs in sleep state also consume energy and thereby generate heat.

3.2 Load Balancing in Cloud Computing UsingDynamic Load Management Algorithm: Ms. ReenaPanwar and Prof. Dr. BhawnaMallick et al.[2] proposed an algorithm that employs Dynamic Load Management taking the set of available virtual machines in a group or block. When a new request comes it checks for best suited virtual machine.Once the request is bound with the virtual machine, this VM index is removed from the group of available virtual machines so it will not be considered for any future request until it finishes its assigned workload and becomes available again by setting its status to be free.If the next upcoming task is received then it checks for the table of VM if it finds a VM, then a request will be assigned to and returns the id of that particular VM to the Data Center, else -1 is obtained. When VM completes its work, the Data Center Controller receives the reply of Datacenter, it notifies the Modified Throttled Load Balancer of the VM de-allocation. The advantage of **t**he proposed algorithm is it gives improved response time because a dynamic set of virtual machines is available and it doesn't consider an overloaded machine again and again for scheduling. Also, it leads to better response time. But, the disadvantage is lack of resource utilization once the VM is allocated with a task it is not considered for next allocation, even if it is lightly loaded.

3.3Energy Efficient Load balancing Algorithm forGreen Cloud: Ms.Tanu Shree and Dr. NeelenderBadal et al.[3] proposed a thermal and power based scheduling policy which will help to reduce the cooling cost and increase the reliability of the computing resources in Cloud Computing environment.

In this paper load balancing or task scheduling technique is proposed which works on power consumption and system temperature. The aim the research work is to reduce the temperature of the computing nodes and to distribute the workload in an efficient way considering thermal power of the system.

The proposed work is divided into two levels.In the first level the cloud administrator whose main function is to create users different types of load (tasks) and systems with different specifications on which the task will work. After generating of the tasks it is allocated to the system with matching specifications.At the second level the scheduler algorithm works as a centralized scheduler collecting all the information about task and dispatching them to the system with appropriate power consumption. Scheduler creates a list of systems having temperature less than threshold value $S_{temp.}$ For each system in S_{temp}queue, power consumption is calculated before allocation of task $P_b$and after allocation of task P_a and del(p)=P_a- P_b is calculated and stored in array S[]. Array S[] is sorted in ascending order and task is allocated to the system with low power value. The advantage of the algorithm is that it avoids overloaded servers and reduces energy consumption, but may increase the response time as it calculates the power consumption before and after the scheduling of each task.

3.4 Predictive Load Balancing in Cloud Computing Environments based on Ensemble Forecasting: Matthias Sommer, Michael Klink, Sven Tomforde, J¨orgH¨ahner et al.[4] proposed a method for the dynamicand anticipatory consolidation of VMs in data centers. Their approach used short-term forecasts of the future utilization of VMs to proactively detect overloaded hosts. In case a host is identified as overloaded, certain VMs have to be migrated to other hosts to avoid SLA violations. The forecasts are generated by a forecast module founded on the idea of ensemble learning using time series of the recent CPU utilization. Based on the analysis of historical utilization data of VMs, forecasts of the future utilization of a host for upcoming time steps can be derived.

If the predicted utilization exceeds a threshold, the host is assumed overloaded. Every 5 minutes (defined

by the scheduling interval), the VM's current CPU utilization in MIPS is calculated as the sum of all processes running onthis VM. This measurement is then passed-on to the forecastmodule. Here, a number of historic CPU utilization valuesis stored in form of a time series.

The Proactive VM migration policy with Utilization Forecasts (PRUF) works as a four-step process:

1) In each time step, the current utilization is monitored and stored. 2) Each VM of host *h* creates a forecast of its utilization for the next migration period. The upcoming utilization of a host is calculated as the sum of all VM utilization forecasts. 3) The utilization percentage of host *h* is then calculated by dividing *host* $mips_{forecast}$by the total number of MIPS of *h*. 4) Finally, a safety parameter $s \in \mathbf{R}$ is applied by utilization$h \cdot s$ to the utilization percentage, e.g. a value of $s = 1.2$ means that the host is assumed overloaded if the predicted utilization exceeds 84% of its available CPU power. In other words, if the utilization percentage exceeds 1, host *h* is regarded overloaded and the VM migration process is triggered, otherwise not.

The benefit of the VM migration policy with proactive overload detection is less performance degradation due to VM migrations and less SLA violations while having only slightly higher energy consumption.

3.5 Improved GA Using Population Reduction For Load Balancing in Cloud computing: Ronak R Patel, Swachil J Patel, Dhaval S Patel, Tushar T Desai et al[5] proposed a GA based on population reduction in cloud computing. The main focus was on balancing the load on most justified resource and also response time to complete the allocated jobs on that. The selection processconsiders original population to define the chromosome, apply Population reduction method to identify best chromosome from set of population, calculate the fitness value with the help of equation F(R) = [LT-TS/LT+TS]+[LT/HL]-[PS/HPS], LT means length of job in MI, PS means Processing speed of Resources in MIPS which is helpful for execution of jobs. HL mean highest length of job from job set & HPS means highest processing speed of resource from resource set. Next step is crossover it's just identified best fit pair of chromosome based on rearranging bits of them and get new set with good and healthy chromosome for future use, now last step is mutation, in that based on toggling of bits from 0 to1 and 1 to 0 a new population is define for next population and new set is ready for further execution. The advantage of this work is that it identified good resources so that resources are able to handle all the upcoming jobs in effective manner which increase overall performance of system. It handles the over-loaded, Underloaded& ideal resources in effective manner so that all resources are utilized properly and give the consistentallotment.

Comparison Table of Various Energy-aware Load Balancing Methods

Load Balancing methods	Resource Utilization	Low Power Consumption	Fast Response Time	Space Overhead	Less Heat Generated
Energy-aware Load Balancing and Application Scaling for the Cloud Ecosystem	**Yes**	**Yes**	**No**	**Yes**	**No**
Load Balancing in Cloud Computing Using Dynamic Load Management Algorithm	**No**	**Yes**	**Yes**	**No**	**No**
Energy Efficient Load balancing Algorithm for Green Cloud	**Yes**	**Yes**	**No**	**No**	**No**
Predictive Load Balancing in Cloud Computing Environments based on Ensemble Forecasting	**Yes**	**No**	**Yes**	**No**	**No**
Improved GA Using Population Reduction For Load Balancing in Cloud computing	**Yes**	**No**	**Yes**	**No**	**No**

4. Conclusion

Energy-efficient managementof virtualizedresources, and in particular dynamic VM consolidation will enable resource providers to successfully offer scalable serviceprovisioning with lower energy requirements and CO2 emissions. Load balancing is key problem in any distributed system thatarea is needs more focus for research. In this paper, we conducted a survey of existing energy-efficient load balancing techniques and discussed different parameters indicating merits and demerits of these algorithms.

References:

1.AshkanPaya and Dan C"Energy-aware Load Balancing and Application Scaling for the Cloud Ecosystem",.MarinescuIEEE Transactions on Cloud Computing

2. Ms. ReenaPanwar and Prof. Dr. BhawnaMallick, "Load Balancing in Cloud Computing Using Dynamic LoadmanagementAlgorithm ":2015 IEEE International Conference on Green Computing and Internet of Things (ICGCloT)

3.Ms.Tanu Shree and Dr. NeelenderBadal "Energy Efficient Load balancing Algorithm for Green Cloud" International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181, Vol. 5 Issue 03, March-2016

4. Matthias Sommer, Michael Klink, Sven Tomforde, J¨orgH¨ahner "Predictive Load Balancing in Cloud Computing Environments based on Ensemble ForecastingOrganic Computing Group", University of Augsburg, Germany, 2016 IEEE International Conference on Autonomic Computing

5. Ronak R Patel, Swachil J Patel, Dhaval S Patel, Tushar T Desai " ImprovedGa Using Population Reduction For Load Balancing In Cloud Computing " 2016 Intl. Conference on Advances in Computing, Communications and Informatics (ICACCI), Sept. 21-24, 2016, Jaipur, India

6. ShridharG.Domanal, G.RamMohana Reddy "Load Balancing in Cloud Computing using Modified Throttled Algoritham" : Department of Information Technology National Institute of Technology Karnataka Surathkal, Mangalore,India

7. AmmarRayes, BechirHamdaoui, MehiarDabbagh and Mohsen Guizaniy, " Energy-Efficient Resource Allocation and Provisioning Framework for Cloud Data Centers", IEEE 2015.

8. Pavithra.B, Ranjana.R , "Energy Efficient Resource Provisioning with Dynamic VM Placement Using Energy Aware Load Balancer in Cloud".