International Journal of Recent Engineering Research and Development (IJRERD)
Volume No. 01 – Issue No. 05, ISSN: 2455-8761
www.ijrerd.com, PP. 15-20

Survey on Energy Aware Load Balancing in Cloud Environment

M. Deepthi[1], N.Senthamarai, M.E, (PH.D)[2]
[1] *PG Student,*
Easwari Engineering College, Tamil Nadu ,India
[2] *Assistant Professor(SR.GR),*
Easwari Engineering College, Tamil Nadu ,India

Abstract: Cloud computing is a new emerging trend in IT environment. It ,is also referred to as "the cloud," is the delivery of on-demand computing resources through the Internet on a pay-as-you-use basis. Load balancing is one of the main challenges in cloud computing .This load balancing requires large amount of energy for distributing the workload evenly across all the nodes. In this paper, we design a heuristic method of load balancing using VM Live migration that efficiently balances the load by migrating VM's to the underutilized nodes rather than the overloaded nodes .Thus my proposed system is (1) To construct a resource usage profile of VMs and hosts , (2) To design load balancing techniques for initial allocation of hosts which may prevent un necessary VM migrations.

Keywords: Cloud Computing, Load Balancing , VM migration, Energy efficient consolidation.

1. Introduction

Cloud Computing is a new Emerging Technology which provides convenient ,on-demand resources from a shared pool to a consumer and releases it later with minimal management effort or interaction of service provider. This model of cloud has 5 characteristics, 3 service models and 4 deployment models.

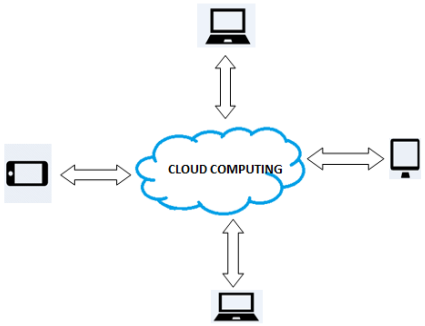

Figure 1: Cloud Computing

1.1 The characteristics in cloud computing are

1.1.1 On-demand self-service is that a consumer is provisioned whatever he needs automatically without requiring human interaction with each service provider.

1.1.2. Broad network access is that a broad network access is required for accessing the resources in cloud.

1.1.3 Resource pooling is The provider's computing resources are placed in a shared pool to serve multiple consumers using a multi-tenant model, from which the resources are fetched and assigned as per the consumer's need.

1.1.4 Rapid elasticity is Capabilities can be elastically provisioned and released automatically, to scale rapidly inward and outward commensurate with demand.

1.1.5 Measured service is Cloud systems automatically control optimize resources and control them by leveraging a metering capability appropriate to the type of service

1.2. The Service Models in cloud computing are

1. **Software as a Service (SaaS**)
2. **Platform as a Service (PaaS)**
3. **Infrastructure as a Service (IaaS).**

1.3. The Deployment Models in cloud computing are

1.3.1 Private cloud: The infrastructure is provided for the use of a single organization consisting of multiple consumers. It may be operated, owned and managed by a third party, organisation or both of them, and it may exist off or on premises.

1.3.2 Community cloud: The infrastructure is provided for the use of a specific community of consumers that have shared concerns. It may be operated , owned and managed by one or more of the organizations in the community, a third party, or some combination of them, and it may exist off or on premises.

1.3.3 Public cloud: The infrastructure is provided for open use by the general public. It may be operated , owned and managed by a academic, business or government organization, or combination of them.

1.3.4 Hybrid cloud: The infrastructure consists of two or more cloud infrastructures such as private, public or community, but are bound together technology that enables data and application portability .

2. LOAD BALANCING

Load Balancing in an important concept in cloud computing, is the process of distributing workloads across computing resources. This Cloud load balancing reduces costs and maximizes availability of resources.

3. IMPORTANCE OF LOAD BALANCING

Cloud Computing has advantages such as Flexibility , Cost and availability of service users .These advantages are the reason that cloud services has a great demand. But this demand leads to many issues in Internet of services style applications and service oriented Architecture like High availability and Scalability. As a result , Load Balancing allows to "scale up to the demands" by allocating workloads evenly to all hosts efficiently.

4. LIVE MIGRATION

Live migration refers to the process of moving a running virtual machine or application between two different hosts or machines without disconnecting the client or application. Network, Memory and storage connectivity of the virtual machine are transferred from the original guest machine to the destination machine.

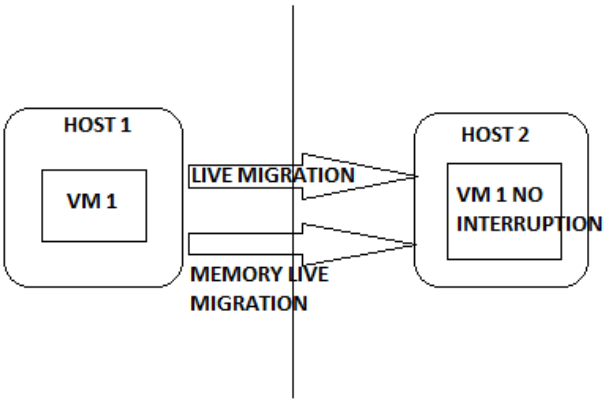

Figure 2: VM Live Migration

Using virtual machine migration achieves:
a) Load balancing among the resources by distributing overall workload.
b) Minimize energy consumption by optimizing resource utilization of resources.
c) Improve the availability of the resources and avoid failures.

Using a load balancing migration technique, we can distribute load across the physical servers in a data center. Dynamic workload in cloud data center manages by live migration of VMs. The Load balancing with VM migration gives benefits like (1) improve resource utilization, (2) enhancing scalability, (3) avoiding bottlenecks, (4)avoiding over provisioning of resources etc.

Two techniques are there for moving the VM's memory state from the source to the destination are

(1) pre-copy memory migration and

(2) post-copy memory migration.

4.1 PRE-COPY MEMORY MIGRATION

This type of migration has two phases.

4.1.1 Warm-up phase

While the VM is actively running on the source, all the memory pages from source to destination are copied by the hypervisor during the pre-copy memory migration. Some memory pages that are changed during this process will be re-copied till the rate of re-copied pages is not less than page dirtying rate.

4.1.2 Stop-and-copy phase

After the warm-up phase, the VM is stopped on the original host and this VM is resumed on the destination host, then the remaining dirty pages are copied to the destination. Down-time is the time between stopping the VM on the original host and resuming it on destination , and it ranges from a few milliseconds to seconds depending on the size of memory and applications running on the VM.

4.2. POST-COPY MEMORY MIGRATION

Suspending the VM at the source initiates the Post-copy VM migration process. With the VM suspended, a minimal subset of the execution state of the VM (CPU state, registers and, optionally, non-pageable memory) is transferred to the target and then the VM is resumed at the target. Pre-paging occurs concurrently where the source actively pushes the remaining memory pages of the VM to the target. Page-fault is generated at the target, if the VM tries to access a page that has not yet been transferred. These are network faults and are trapped at the target and redirected to the source, which responds with the faulted page.

4.3 SEAMLESS LIVE MIGRATION

A live migration where the down-time of a VM is not noticeable by the end user is called as a seamless live migration.

5. NEED FOR ENERGY AWARE CONSOLIDATION IN CLOUD COMPUTING

The growth and demand of cloud computing has lead to increased in the Energy Consumption of data centers . This has become a major issue for both society and industry. This increase in energy leads to increase in cost. Another reason for energy consumption is that the servers are always ON. Thus there should be solution to minimize the energy consumption which results in profit margin for the users. Thus this is the need for energy consolidation in cloud computing.

6. RELATED WORK ON ENERGY-AWARE CONSOLIDATION:

6.1. C AND E PROTOCOL: J. Octavio Gutierrez-Garcia et al.[9] presented that Load management in cloud data centres need high resources, high energy consumption and high requirements .

So this paper proposed distributed problem solving techniques in cloud data centres using the use of Collaborative agents .These agents collaborate with each other to perform VM migrations using two protocols (1) Collaborative load management protocol, (2) Energy-aware server consolidation protocol .Thus the overload achieved is lower than the overhead of centralized and other distributed approaches .But this paper has issues such that (1)It is suitable only for single organisation (2) Statistical resource usage of VM’s and hosts are not explicit.

6.2. ENERGY CONSERVATION TECHNIQUES: Mehiar Dabbagh et al.[10] investigated that Energy consumption has become a big problem in cloud data centres. Thus many innovative ideas are being proposed for less energy consumption .In this paper,

Three Energy Conservation Techniques are used (1) Workload prediction (2) VM Placement and workload consolidation (3) Resource overcommitment .Thus, great energy is saved by turning more servers into power mode and by increasing the utilization of already active ones.

6.3. MINIMIZATION OF MIGRATIONS: Anton Beloglazov et al.[2] investigated that Data centres hosting cloud computing may need high energy consumption , high operational costs and more carbon foot prints in

environment. Thus Green computing should be used to reduce the operational cost, energy consumption and to reduce the carbon emissions .

In this paper, many energy consumption ideas are surveyed using two methods (1) minimization of migrations (2) Modified Best Fit Decreasing .This approach leads to substantial reduction of energy consumption which will give way to Green computing when compared to static resource allocation techniques. It has also proposed that cost of software engineering can be reduced by reusing the existing cloud middleware and its technologies.

6.4. FIREFLY APPROACH: Nidhi Jain Kansal et al. [11] presented that Energy efficiency has become latest growth in the area of cloud computing. The increased growth in the number and size of data centres have lead to energy conservation. This energy can be conserved only by reduced VM live migrations.

In this paper , a firefly algorithm based approach is used which migrates maximal overloaded hosts to least loaded active nodes which reduces the number of migrations.This is done by two algorithms (1)Firefly Optimization (FFO) Algorithm (2) FireFly Optimization—Energy-Aware Virtual Machine Migration (FFO-EVMM)Technique

Thus energy is conserved which contributes towards Green computing. Thus energy consumption has attained 44.39% by reducing number of migrations to 72.34% and saving 34.36% of hosts.

6.5. **LEEDS TEST BED APPROACH**: Ibrahim Alzamil et al.[6] presented a paper that Cloud computing has changed the way how people use resources nowadays. We can use the cloud computing resources instead of buying it using a pay-per-use concept.Since cloud computing is highly used nowadays ,the cost for maintaining the infrastructure of cloud is very much important. In this paper. A system architecture is proposed for profiling the resource usage in terms of energy consumption. From this data, the developers should be able to enhance and optimize the energy consumptions.here leeds test bed is used to support energy – awareness.

Thus, the result shows that Leeds test bed is installed and it is used to support energy-awareness at both physical hosts and The conducted experiments showed that Leeds testbed has been setup as a Cloud environment supporting energy-awareness at both
physical host and VM levels.

6.6. SLA VIOLATION DECISION ALGORITHM:Zhibo Cao et al.[14] did research on virtual machine (VM) consolidation in cloud computing.This consolidation provides a great opportunity of energy savings.

In this paper, they have two algorithms (1) Service Level Agreement (SLA) violation decision algorithm (2) minimum power and maximum utilization policy.

SLA violation algorithm is to check whether a host is overloaded with SLA violation. Min power max utilization policy is to improve the minimum power policy.Thus, the simulatiuon result shows that there is 21-34% decrease in energy consumption, 84-92% decrease in SLA violation, 87–94 % decrease in energy-performance metric, and 63 % decrease in execution time.

6.7. PREANT POLICY APPROACH: Hancong Duan et al.[4] investigated how to reduce energy consumption and to increase the computational capacity. Existing schemes focussed to increase resource utilization using "bin-packing" algorithm. In this paper, authors proposed a novel scheduling approach called "PreAntPolicy" which is a prediction model and a scheduler of an Ant algorithm. This prediction model decides whether to trigger the execution of scheduler .The scheduler which does resource scheduling minimizes the energy consumption based on the QOS factors.

Thus , this approach had reduced energy usage upto 75.28 % for round robin and to 76.17% for first fit. This scheduler also works for multiple resource scheduling in a real world scenario. Thus ,there is a excellent energy efficiency as well as resource utilization. It also had good results of energy efficiency in high peak loads.

6.8. MODIFIED SWARM OPTIMIZATION METHOD: Hongjian Li et al.[5] investigated about energy efficient VM migration and consolidation algorithm to reduce energy consumption based on QOS factors. In this paper, the authors designed a double threshold method to trigger the migrations in a multi resource utilization.Thus to avoid local optima, which is a common defect in traditional algorithms, the modified particle Swarm Optimization method is used for VM consolidation.

Thus the analysis between traditional algorithms showed that our algorithm has improved in energy efficiency ,reduced the migrations as well as active physical nodes and balanced the resources.

6.9. AUTOMATIC SERVER CONSOLIDATION: Paul J. Kuehn et al.[12] investigated about Data centre's operations in terms of cost, energy consumption and cooling requirements. Now, many efforts are made for

server virtualization and consolidation in order to get a proportionality between energy consumption and computation amount.

In this paper,(1) a model for automatic server consolidation by load dependent control of server using hysteresis thresholds,(2) (DVFS)Dynamic Voltage and Frequency Scaling and (3) hot and cold server standby .A multi server queuing model controlled by Finite State Machine(FSM) is used for energy efficiency and analysis of performance .This model should accept for QOS parameters except overload conditions.
This model also provides a study of trade-off between performance and power efficiency .This model gets exact meaning only when used with Markovian assumptions .The proposed model is to include many parameters , such as power saving mode, load dependent activation/deactivation thresholds.

6.10. GEOGRAPHIC LOAD BALANCING ALGORITHM: Debdeep Paul et al.[3] presented a paper based on cloud service providers. This cloud service providers(CSP) has multiple –geographically distributed data centres with renewable energy resources.They also did a research based on Load distribution to reduce electricity cost, accept to SLA agreement based on the effect of server switchings.
Their work provides details about performance of different algorithms for Geographic Load Balancing(GLB) in terms of cost and renewable energy, for this they proposed two MPC based algorithms .Their proposed work is also to include way of capturing server switching costs .The inputs to the system are (1) electricity price (2) renewable energy (3) number of job requests may vary which will lead to fluctuations in the average cost of electricity per job request .These fluctuation can be controlled by leveraging contracts of thgis average cost per job request in the forward electricity market and to determine the cost of electricity in that market. The results show that there is a reduction in the average cost of electricity per job request as well as the price risk mitigation is achieved with reduction in cumulative cost of electricity. Thus the reduction in cumulative electricity cost will result in considerable amount of cost savings.

Table 1:

TECHNIQUES	METRICS
C and E protocol	Overload achieved is lower
Energy Conservation Techniques	Great energy savings achieved
Minimization of Migrations	Substantial reduction of energy consumption
Firefly approach	Reduces energy consumption, Higher Scalability
Leeds Testbed approach	Profiling and assessing the energy efficiency
SLA Violation Decision Algorithm	Decrease in energy consumption, SLA violation, Energy performance and execution time
PreAnt policy approach	Energy aware scheduling for virtual machines
Modified Swarm Optimization method	Improved energy efficiency
Automatic Server Consolidation	Values of parameters are maintained
Geographic Load Balancing Algorithm	Reduction in average cost of electricity per job request and cumulative cost of electricity

7. References

[1]. Aarti Singha , Dimple Junejab , Manisha Malhotraa ,"Autonomous Agent Based Load Balancing Algorithm in Cloud Computing",ELSEVIER, Procedia Computer Science ,Volume 45,Pages 832-841, March 2015.

[2]. Anton Beloglazov, Jemal Abawajy , Rajkumar Buyya," Energy-aware resource allocation heuristics for efficient management of data centers for Cloud computing",ELSEVIER, Future Generation Computer Systems**,**Volume 28,Issue 5, May 2012, Pages 755–768

[3]. SanjayK.Bose," Energy efficiency aware load distribution and electricity cost volatility control for cloud service providers",ELSEVIER,Journal of Network and Computer Applications,Volume 59,pages 185-197,January 2016.

[4]. Hancong Duan , Chao Chen , Geyong Min , Yu Wu ,Elsevier," Energy aware scheduling of virtual machines in heterogeneous cloud computing systems",ELSEVIER,Future Generation computer Systems,March 3,2016

[5]. Hongjian Li ,Guofeng Zhu ,Chengyuan Cui, Hong Tang, Yusheng Dou ,Chen He," Energy-efficient migration and consolidation algorithm of virtual machines in data centers for cloud computing",Springer,Computing,Volume 98,Issue 3,pp 303-317,March 2016.

[6]. Ibrahim Alzamil, Karim Djemame, Django Armstrong and Richard Kavanagh, ," Energy-Aware Profiling for Cloud Computing Environments",ELSEVIER, Electronic Notes in Theoretical Computer Science, Volume 318, Pages 91-108 , 25 November 2015 in Twenty-ninth and thirtieth Annual UK Performance Engineering Workshops (UKPEW).

[7]. Javier Bajo , Fernando De la Prietaa, Juan M. Corchadoa, Sara Rodríguez," A low-level resource allocation in an agent-based Cloud Computing platform", ELSEVIER, Applied Soft Computing,Volume 48, November 2016, Pages 716–728.

[8]. J. Octavio Gutierrez-Garcia1 · Adrian Ramirez-Nafarrate,"Agent-based load balancing in Cloud data centers",Springer, Cluster Computing, Volume 18,Issue 3, pp 1041-1062, September 2015.

[9]. J. Octavio Gutierrez-Garcia and Adrian Ramirez-Nafarrate ,"Collaborative Agents for Distributed Load Management in Cloud Data Centers Using Live Migration of Virtual Machines" ,IEEE transactions on Services Computing , Volume:8,Issue:6,November/December 1,2015.

[10]. Mehiar Dabbagh , B.Hamdaoui , M.Guizani and A.Rayes, "Toward energy efficent cloud computing prediction , consolidation and over commitment",IEEE Network,Volume 29,Issue 2, March/April 2015.

[11]. Nidhi Jain Kansal ·Inderveer Chana ,"Energy-aware Virtual Machine Migration for Cloud Computing - A Firefly Optimization Approach",Springer, Journal of Grid Computing, Volume 14,Issue 2,Pp 327-345, June 2016.

[12]. Paul J. Kuehn, Maggie Ezzat Mashaly, "Automatic energy efficiency management of data center resources by load-dependent server activation and sleep modes", ELSEVIER, Ad Hoc Networks ,Volume 25,part B, February 2015, Pages 497–504

[13]. Young Choon Lee, Albert Y. Zomaya," Energy efficient utilization of resources in cloud computing systems"Springer,The Journal of Supercomputing,Volume 60,Issue 2 ,pp 268-280,May 2012.

[14]. Zhibo Cao · Shoubin Dong," An energy-aware heuristic framework for virtual machine consolidation in Cloud computing",Springer ,the Journal of Supercomputing,Volume 69,Issue 1, pp 429-451,July 2014.

[15]. Zhuo Tang,Ling Qi,Zhenzhen Cheng,Kenli Li,Samee U.Khan , "An Energy-Efficient Task Scheduling Algorithm in DVFS-enabled Cloud Environment",Springer,Journal of grid Computing ,Volume 14,Issue 1,pp 55-74,March 2016.