Cloud Computing Load Balancing Model with Heterogeneous Partition for Public Cloud

Ms. Pranita Narayandas Laddhad[1] , Prof. Nitin Raut[2], Prof. Shyam P. Dubey[3]
M. Tech. (2nd Year) CSE, Nuva college of Engg. & Tech., Nagpur, India [1],
Asst. Professor, Dept. Of CSE, GNIET, Nagpur, India[2]
Asst. Professor, Dept. Of CSE, Nuva college of Engg. & Tech., Nagpur, India[3]
prnita.laddhad@gmail.com[1], infotech.nitin@gmail.com[2], shyam.nuva@rediffmail.com[3]

ABSTRACT

Cloud computing is a latest feature of widespread distributed computing. In the cloud computing paradigm, the scheduling of computing resources is a difficult part of cloud computing field. With increment in number of users and the type of applications on the cloud computing, effective utilization of resources in the system becomes a critical concern. In order to improve the performance of the whole cloud environment, Load Balancing algorithms are needed to distribute the load evenly across all the nodes in cloud. In this system Cloud having numbers of nodes and that is distributed in many partitions according to need. System having clouds each cloud having different capacity. In this system admin can set the size of cloud dynamically, it can change the size of any cloud according to need. Admin upload the jobs on to cloud then it redirect the job on different cloud according to load degree calculation cloud status will change from idle to normal and from normal to overloaded if there is no free load available on that cloud. When partition is overloaded at that time load is distributed over the other ideal nodes. Many algorithms are available for load balancing like Static load balancing and Dynamic load balancing. System used round robin with load degree evaluation algorithm which is dynamic load balancing algorithm. As job is redirected towards the cloud, cloud active status will increase. If all partition in the system is overloaded then system gives message maximum connection are crossed. Users not allowed upload the file but he can view the file directly without authentication. The load balancing criteria have significant influence on the performance, and they can change the behavior of the system. The load balancing model given in this article is aimed at the cloud which has numerous nodes with distributed computing resources in many different geographic locations. Thus, this model divides the cloud into several cloud heterogeneous partitions. When the environment is very large and complex, these divisions simplify the load balancing. The cloud has a main controller that chooses the suitable partitions for arriving jobs while the balancer for each cloud partition chooses the best load balancing strategy. The Round Robin algorithm is used here because it is fairly simple.

KEYWORDS: Heterogeneous Cloud Partition, Round robin with load degree evaluation, Load Balancer, Public Cloud.

INTRODUCTION

Cloud computing is an on demand service in which shared resources, information, software and other devices are the clients requirement at specific time. It's a term which is generally used in case of Internet.

The whole Internet can be viewed as a cloud. Capital and operational costs can be cut using cloud computing. In case of Cloud computing services can be used from diverse and widespread resources, rather than remote servers or local machines. There is no standard definition of Cloud computing. Generally it consists of a bunch of distributed servers known as masters, providing demanded services and resources to different clients known as clients in a network with scalability and reliability of datacenter. The distributed computers provide on-demand services. Services may be of software resources (e.g. Software as a Service, SaaS) or physical resources (e.g. Platform as a Service, PaaS) or hardware/infrastructure (e.g. Hardware as a Service, HaaS or Infrastructure as a Service, IaaS). Amazon EC2 (Amazon Elastic Compute Cloud) is an example of cloud computing services. In a distributed system, dynamic load balancing can be done in two different ways: distributed and non-distributed. In the distributed one, the dynamic load balancing algorithm is executed by all nodes present in the system and the task of load

balancing is shared among them. Dynamic load balancing algorithms of distributed nature, usually generate more messages than the non-distributed ones because, each of the nodes in the system needs to interact with every other node. Distributed dynamic load balancing can introduce immense stress on a system in which each node needs to interchange status information with every other node in the system. It is more advantageous when most of the nodes act individually with very few interactions with others.

TECHNIQUES LITERATURE

In literature, we study most of recent load balancing algorithm that have been developed in cloud computing domain. Many different static and dynamic algorithm is developed for proper load balancing. Dhinesh et al. [5] proposed an algorithm named honeybee behavior inspired load balancing algorithm. Here in this session well load balance across the virtual machines for maximizing the throughput. This algorithm is derived from the behavior of honey bees that uses the method to find and reap food. The tasks removed from the overloaded VMs act as Honey Bees. VMs executes the high priority process first. Zhang et al.[8] proposed a binary tree structure that is used to partition the simulation region into sub-domains . The characteristics of this fast adaptive balancing method are to be adjusted the workload between the processors from local areas to global areas. According to the difference of workload, the arrangements of the cells are obtained. Yunhua.et al.[9] proposed an efficient cell selection scheme and two heat diffusion based algorithm called global and local diffusion. Considered the distributed virtual environments there were various number of users and the load accessing by the concurrent users can cause problem. This can be avoided by this algorithm. According to the heat diffusion algorithm, the virtual environment is divided in to large number of square cells and each square cells having objects.

ARCHITECTURE OF MODEL

System provides the load balancing in cloud computing for heterogeneous partition. Every partition in this system having different no of nodes according to need. If location having more traffic then on that partition nodes are more than other, For region having less traffic partition having minimum node. Admin login the system after authentication , check all cloud status if all cloud is at idle condition then start uploading the file and then redirect to cloud1 and so on. According to that partition status will change from idle to normal and then from normal to overloaded. A cloud partition has several nodes belongs to a particular area, these subarea of the public cloud based on the geographic locations.

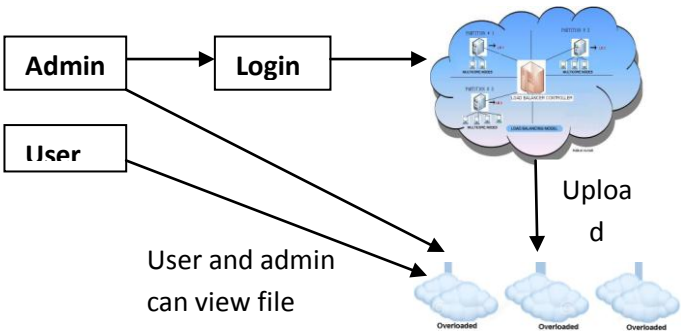


Fig1. Architecture of load balancing model based on cloud computing for heterogeneous partition

There is a main controller system which manages the load called Load Balancer Manager (LBM). After partitioning the public cloud into different partitions, load balancing is starts. Then assign job will send on to cloud and simultaneously status of cloud will change from idle to normal and from normal to overloaded respectively.

Working of LBM

In this model, Load Balancer Manager (LBM) is responsible for the following task-

- Receives the jobs from different users one by one.
- Choose a specific partition for the received jobs one by one start from cloud1.
- Check the status of the cloud partition (Status may be in one of these: IDLE, NORMAL, and OVERLOADED).
- If the partition Status= OVERLOADED the no allocation will be done, it means all nodes are overloaded already.

METHODOLOGY USED

Round Robin based on the load degree:

In this algorithm, Each process is assigned to the processor in a round robin order. The process allocation order is maintained locally independent of the allocations from remote processors. Though the work load distributions between processors are equal but the job processing time for different processes are not same. . A load degree is assigned to each server based on criteria chosen by the site administrator; the most commonly used criterion is the server's traffic-handling capacity. The higher the weight, the larger the proportion of client requests the server receives. If, for example, server B having Idle status and server A having normal status, the load balancer forwards incoming requests to server B than it sends to server A. This algorithm works well when you have two or more Cloud Servers that are unequal in computing power and available resources. The main benefit of round-robin load balancing is that it is extremely simple to implement.

SYSTEM MODULE

System give the load balancing strategy is based on cloud partition concept. Partition of cloud is decided before stating the load balancing. According to partition admin decide which cloud receive how many jobs according to their capacity.

A. Cloud Partition

Cloud partition in load balancing makes cloud computing more efficient and improves user satisfaction. Cloud is partition into many different partition it include many nodes. System gives heterogeneous partition of cloud i.e. if partition A having more load then assign greater no of nodes to partition A. System Assign 2 nodes to Partition 1 , 3 nodes to partition 2 and 4 nodes to partition 3, respectively. The cloud partition Model is shown in Fig 2

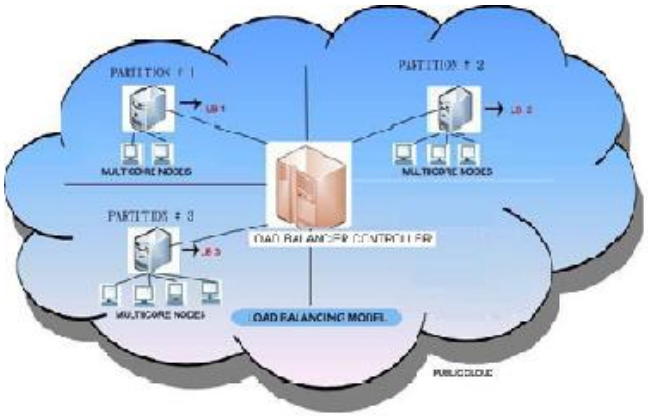


Fig2. Cloud Partition Model

B. Assigning jobs to the cloud partition

For assigning Job to partition the Round Robin with load degree evaluation algorithm is used. If all partition having idle status then first job is assign to partition 1 and when first partition change the status from idle to normal according to strategy then next job assign to partition 2 and so on.

Algorithm-1 Best Partition Searching

Begin
While job do searchBestPartition (job);

Select 1st partition having lowest status value

If partitionState == idle

then Send Job to Partition;

else if partitionState == normal

then Send Job to Partition;

Else

Maximum connection is crossed;

if End

While end

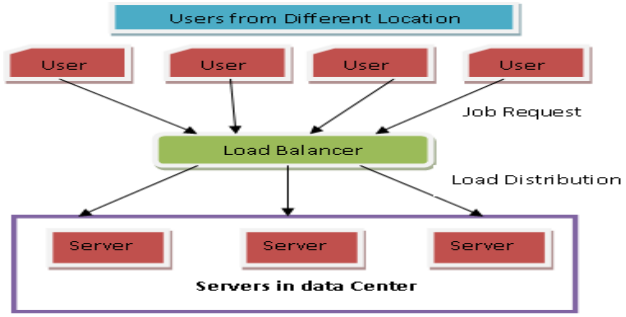
C. Workload balance strategy for the idle status, normal status and overloaded status.

When all cloud partition having idle status, many computing resources are available and relatively few jobs are arriving. In this situation, this cloud partition has the ability to process jobs as quickly as possible so a simple load balancing method can be used. When partition nodes reach to half of total nodes then partition status change from idle to normal. When partition capacity is fulfilled then status change from to overloaded. After the job is uploaded then system decide to redirect the job towards the cloud. The status of the partition is change according to load degree

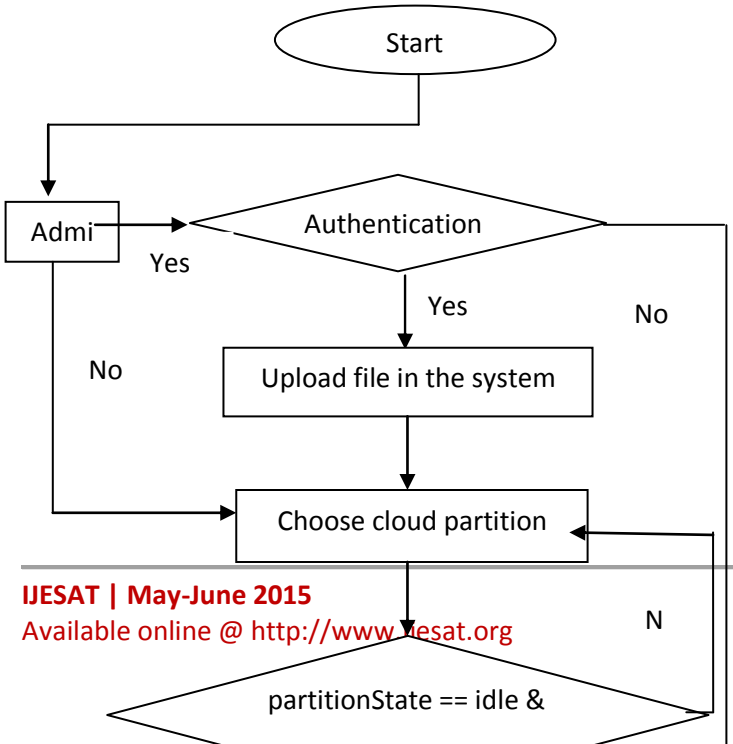
- Idle When Load degree.(N) == 0 Or Load degree.(N) < ½ (Total Nodes);
  There is no job being processed or some jobs are there but not exceed idle state limit so status is idle
- Normal For (Total Nodes) < Load degree.(N) => ½ (Total Nodes);
  The node is normal and it can process other jobs
- Overloaded When Load degree(N) == (Total Nodes);
  Now There is no space to processed more job.

IMPLEMENTATION

Today the whole environment is more complex and huge, the load balancing technique is used to simplify the environment by these division. Then suitable partition can be select by admin for after file is uploaded However, the balancer chooses the best suitable load balancing strategy. In the past system the size of the cloud partition is same i.e. the server is divided into partition having equal numbers of nodes because of this the location having greater demand and location having less demand is considered as equal priority and In previous many load balancing system server has not been set and also it is not taken as a parameter while uploading the load to the server so there might be a risk of overflowing the server The system will check the total space and the actual size of the server vary from server to server. After uploading the file it will check the size of file and the file will be uploaded to the other server, if capacity of partition is overloaded then we can't redirect more file on that partition.



Flow Diagram



MODULE ANALYSIS

A. Admin Collaboration in System

A public cloud is one based on the standard cloud computing model, in which a service provider makes resources, such as applications and storage, available to the general public over the Internet. User uses the service provided by public cloud. After admin login successfully, file is uploaded by admin and then that file or job is redirect to clouds. Before job redirection all cloud having idle status. The status of partition change according to load_degree(LD). Idle status change to normal when load_degree is between zero and high_LD and then status change to overload after load_degree is reach at high_LD level. When LD reached the overloaded status then admin is not

allowed to redirect the job on to that partition, systems choose another partition for job redirection. Then Admin and User can view the jobs directly without login or without registration on public cloud.

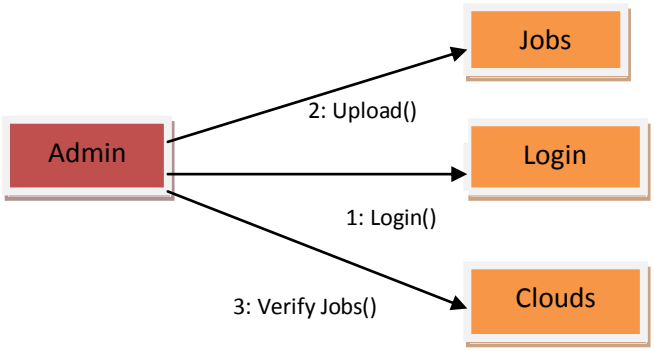


Fig4. Admin Collaboration in the System

B. Uploaded the Jobs in the system and redirects it to Cloud Partition

After login admin is permitted to upload the jobs in the system. There are three fields in this model File Name, Category, and Choose job from System then upload the file on the cloud. After job is uploaded on the cloud , first step is to choose the cloud partition among three , at starting it choose first cloud. Cloud partition can be in the following status:

IDLE: If partition having more than haft connection is remaining.

NORMAL: If partition having more than half of total connection but not overloaded.

OVERLOADED: If Maximum connection is crossed.

Any cloud partition having the status= OVERLOADED is not selected by the Load Balancer Manager and likewise any node having the Load Degree (LD) =OVERLOADED is not eligible for the processing. Only cloud partition having IDLE or NORMAL load status and Node having IDLE or NORMAL load degree are selected for scheduling. Compute the Load Degree (LD) of a node in any cloud partition is as follows :

$$LD\ (CN) = \sum_{i=1}^{m} Xi * Pi$$

Here, CN=Current Node, Pi are the parameter either static or dynamic where Pi(1<=i<=m), m represents the total number of parameter . Xi are weights that may differ for different kinds of job for all (1<=i<=n). Average Load Degree (LD) of the cloud partition will be calculated as-

$$Avg\_LD = \sum_{i=1}^{n} LD\ (CNi)/n$$

The Sequence for uploading the file and redirect it to cloud partiotn is given in figure5

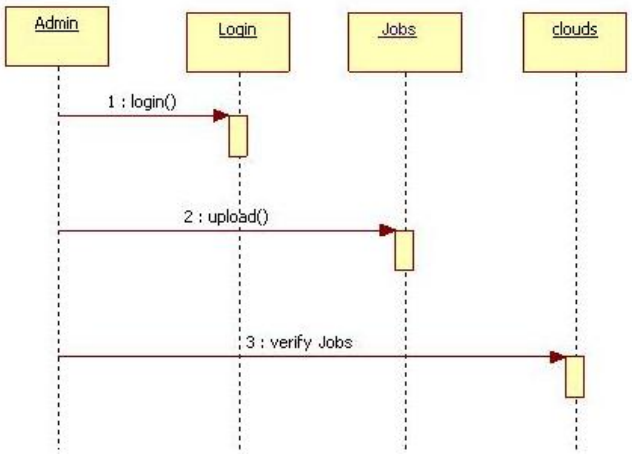


Fig5. Upload the file and redirect towards Cloud

RESULT ANALYSIS & DISCUSSION

Paper shows how load is balance amount all cloud partition which is heterogeneous in size. Project start at login page, only authenticated admin can login. We can set or change the size of cloud or no of nodes per cloud by using cloud size webpage as shown in fig.6

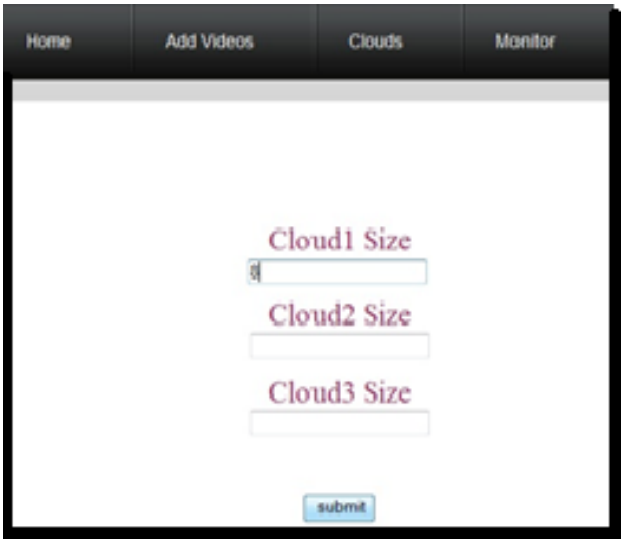


Fig.6 Cloud Size on Server

As job is uploaded on server the suitable cloud partition is seleced according to calculation on load degree. Uploading of job on cloud as shown in fig.7

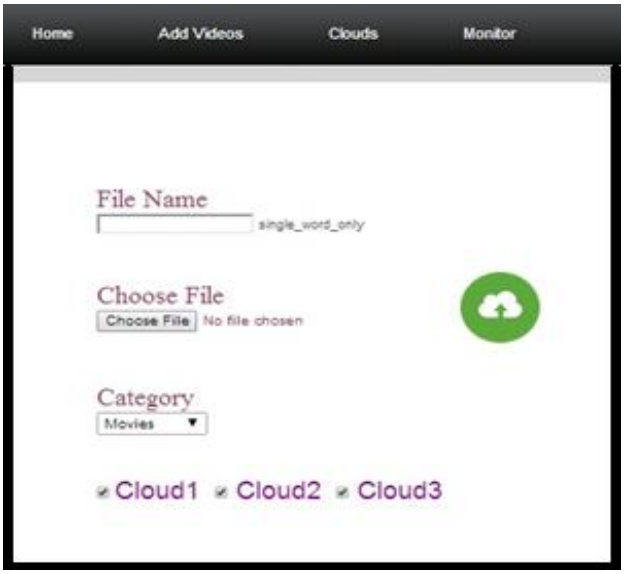
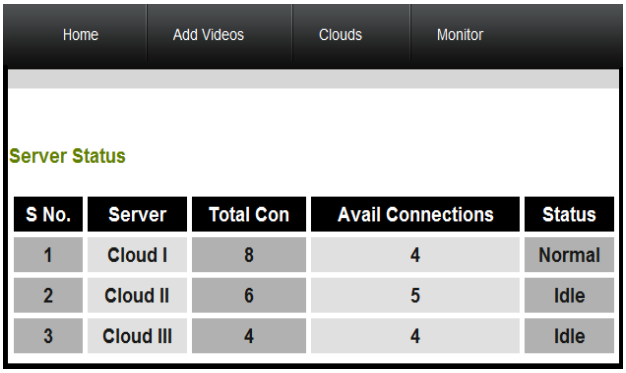


Fig.7 Upload the Job on Server

At the stating Position all cloud Serer having Idle Status. As job is uploaded the server status change from Idle to Normal and if all avail connection is reached to zero then status changed to overloaded. The Server status is shown in figure 8.

Simultaneously Active status of cloud is shown in figure 9.



S No.	Server	Total Con	Avail Connections	Status
1	Cloud I	8	4	Normal
2	Cloud II	6	5	Idle
3	Cloud III	4	4	Idle

Fig.8 Server status of Cloud after redirecting Jobs

Fig.9 Active status of Cloud after redirecting Jobs

CONCLUSION AND FUTURE SCOPE

In this Paper the main aim was to develop jobs balancing model which improved the performance and utilized the cloud resources properly. Cloud is made up of several nodes situated in deferent geographical location. If one location having more requirement then system assign more node on that location. Cloud partition is a method which partitions the total cloud and makes the cloud heterogeneous partition according to location need. This objective is archived by implanting Round Robin with load degree evaluation algorithm in System. After Uploading files by admin, that job is redirected on cloud partition, simultaneously cloud status change. User can view the uploaded jobs on public cloud. After total nodes are assign job we will get maximum connection crossed message. In this way this strategy to improve the efficiency in cloud environment. Future Scope is How to calculate Load Degree: In future provide better load status evaluation, how to set Load degree high and Load degree low for this algorithm is needed. Provide Other load balancing strategy : Developed algorithm which provide better load balancing , to compare system needed other algorithm tests. Many tests are needed to guarantee system availability and efficiency.

REFERENCE

[1] Gaochao Xu, Junjie Pang, and Xiaodong Fu "A Load Balancing Model Based on Cloud Partitioning for the Public Cloud" TSINGHUA SCIENCE AND TECHNOLOGY Volume 18, pp34-391, February 2013.

[2] K. Nishant, P. Sharma, V. Krishna, C. Gupta, K. P. Singh, N. Nitin, and R. Rastogi, Load balancing of nodes in cloud using ant colony optimization, in Proc. 14th International Conference on Computer Modelling and Simulation (UKSim), Cambridgeshire, United Kingdom, Mar. 2012, pp. 28-30.

[3] J. Hu, J. Gu, G. Sun, and T. Zhao, "A Scheduling Strategy on Load Balancing of Virtual Machine Resources in Cloud Computing Environment", Third International Symposium on Parallel Architectures, Algorithms and Programming(PAAP), 2010, pages 89-96.

[4] Z Chaczko, V Mahadevan, S Aslanzadeh and C Mcdermid, "Availability and Load balancing in Cloud Computing". International Conference on Computer and Software Modeling, vol 14, Singapore. 2011.

[5] Dhinesh Babu L.D, P. VenkataKrishna, "Honey bee behaviour inspired load balancing of tasks in cloud computing environments", Applied Soft Computing 13 (2013) 2292–2303.

[6] Sachin Kumar, Niraj Singhal, "A Study on the Assessment of Load Balancing Algorithms in Grid Based Network", published in the International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-1, March 2012.

[7] Sandeep Sharma, Sarabjit Singh, and Meenakshi Sharma, "Performance Analysis of load Balancing Algorithms", published in the proceedings of World Academy of Science, Engineering and Technology Volume, 28 April 2008, ISSN 2070-3740.

[8] Zhong Xu, Rong Huang,(2009)"Performance Study of Load Balanacing Algorithms in Distributed Web Server Systems", CS213 Parallel and Distributed Processing Project Report.

[9] Yunhua Deng, Rynson W.H. Lau, "Heat diffusion based dynamic load balancing for distributed virtual environments", Proceedings of the17th ACM Symposium on Virtual Reality Software and Technology, ACM, 2010, pp. 203–21.

[10] Zenon Chaczko, Venkatesh Mahadevan, Shahrzad Aslanzadeh and Christopher Mcdermid , "Availability and Load Balancing in Cloud Computing", IPCSIT vol.14 (2011) © (2011) IACSIT Press, Singapore 2011 International Conference on Computer and Software Modeling.

[11] Eduardo Pinheiro, Ricardo Bianchini, Enrique V. Carrera, and Taliver Heath, "Load Balancing and Unbalancing for Power and Performance in Cluster-Based Systems" Technical Report DCS–TR–440, May 2001