ISSN:0975-9646

Semi-Distributed Cloud Computing System with Load Balancing Algorithm

Payal A.Pawade[#], Prof. V. T. Gaikwad[*]

[#]*CSE Department, SGBAU Amravati University*
SIPNA COET Amravati MH
[#] *Associate Professor , CSE Department, SGBAU Amravati University*
SIPNA COET Amravati MH

***Abstract*— Cloud computing is organize in the data centre where physical machine are virtualized can be seen in this paper. Over the Internet in general, Cloud computing is a term used for anything that involves delivering hosted services. As cloud computing is a new technology which has both merits and demerits, load balancing is one of the major issue faced by cloud computing. In Virtualization many virtual machines can be run. Many researches & studies need to be carried out for Load balancing as it is an important topic in cloud computing. The data centre is built with many systems where balancing is becomes a very difficult task especially for cloud computing. Most of the research of cloud computing is done in distributed environments. Distributed load balancing on cloud computing is already in the list; despite of the fact that the use of semi-distributed load balancing in cloud computing is not discussed in any literature. A new algorithm for the cloud computing can be designed by using the method of semi-distributed load balancing.**

***Keywords*- Semi-distributed load balancing, virtualization, physical machine,virtual machines.**

I. INTRODUCTION

Cloud computing is an on demand service in which shared resources, information, software and other devices are provided according to 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.

Load balancing in cloud computing systems is really a challenge now. Always a distributed solution is required. Because it is not always practically feasible or cost efficient to maintain one or more idle services just as to fulfill the required demands. Jobs can't be assigned to appropriate servers and clients individually for efficient load balancing as cloud is a very complex structure and components are present throughout a wide spread area. Here some uncertainty is attached while jobs are assigned.

In a cloud computing environment, computers are connected with completely different physical machine (PM) or host above the PM, several virtual machines (VM) were deployed by using the technology call Virtualization or hypervisor (e.g. Xen, KVM, VMware), VM are configured with different properties (RAM, storage, CPU etc.) On top of the Hypervisor in which Operating System (OS) is installed like normal system. Different apps and service were run to produce an unlimited resource of computing to the end user.

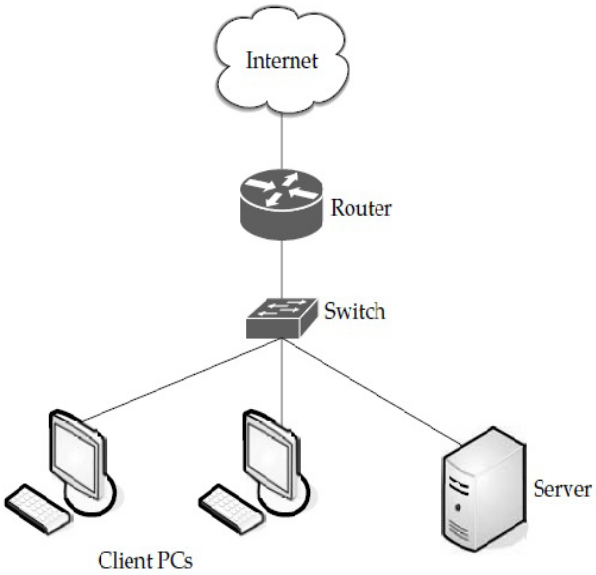

Fig 1: A cloud is used in network

The advantage of using cloud computing is that in a single system we can hypervise different OS to deploy different service like FTP, DB, Web Service and email processing for that in traditional we need different four computers, using of Virtualization technology helps us run to all the service in a single machine. By using Virtualization technology we can run different OS on a single system. By this we can keep two systems one for deploying all the apps and service and one can be kept for online backup. Migration of all VM is possible without the need of same physical machine also if anything happens to the system [1]. The Figure 1 is how the Virtualization is done.Increase/decrease of application traffic can be responded automatically by Load balancing which enables IT to achieve even greater fault tolerance.The load balancer is able to handle half of the traffic in a cloud configuration to fully handle the traffic need to scale the numbers of the servers in the back end mainly for CPU- intensive applications [2]. To work efficiently and smoothly load balancing are implemented on different part of computing. Hardware load balancer and software load balancer are widely used, using of hardware load balancer is quite costly so some prefer to use software load balancer for their solution,open source load balancer which are widely used

are Apache web Server's mod_proxy_balancer extension, Varnish, or the Pound reverse proxy and load balancer [3].

cloud components:

A Cloud system consists of 3 major components such as clients, datacenter, and distributed servers. Each element has a definite purpose and plays a specific role.

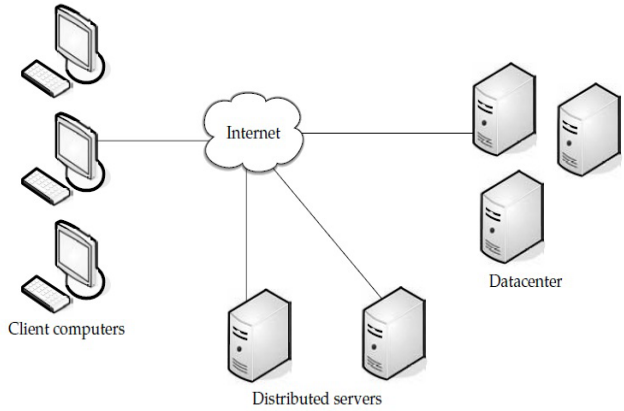

Fig 2: Three components make up a cloud computing solution

Clients

End users interact with the clients to manage information related to the cloud. Clients generally fall into three categories as:

Mobile: Windows Mobile Smartphone, smartphones, like a Blackberry, or an iPhone.

Thin: They don't do any computation work. They only dispaly the information. Servers do all the works for them. Thin clients don't have any internal memory.

Thick: These use different browsers like IE or mozilla Firefox or Google Chrome to connect to the Internet cloud.

Now-a-days thin clients are more popular as compared to other clients because of their low price, security, low consuption of power, less noise, easily replaceble and repairable etc.

Datacenter

Datacenter is nothing but a collection of servers hosting different applications. A end user connects to the datacenter to subscribe different applications. A datacenter may exist at a large distance from the clients. Now-a-days a concept called virtualisation is used to install a software that allow multiple instances of virtual server applications.

Distributed Servers

Distributed servers are the parts of a cloud which are present throughout the Internet hosting different applications. But while using the application from the cloud, the user will feel that he is using this application from its own machine.

II. Review of Literature

Cloud Computing is the most recent topic in IT industry due to its flexibility in using the computing system. Cloud computing provides everything as a service, some of the service which is widely available in the IT market is a SaaS (Software as a service), PaaS (Platform as a service), IaaS (infrastructure as a service) [7] [5].

2.1 Software as a service (SaaS)

In SaaS, the user uses different software applications from different servers through the Internet. The user uses the software as it is without any change and do not needs to make lots of changes or doesn't require integration to other systems. The provider does all the upgrades and patching while keeping the infrastructure running [2].

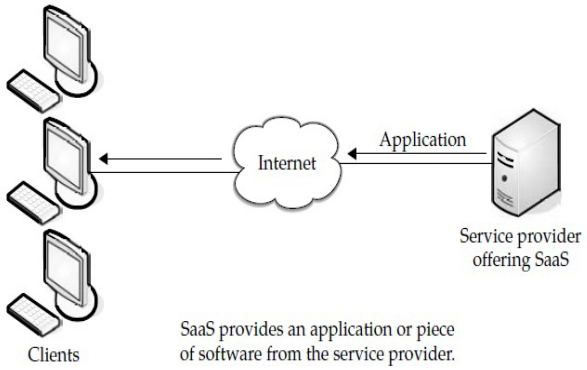

Fig 3: Software as a service (SaaS)

The client will have to pay for the time he uses the software. The software that does a simple task without any need to interact with other systems makes it an ideal candidate for Software as a Service. Customer who isn't inclined to perform software development but needs high-powered applications can also be benefitted from SaaS.

Some of these applications include:

Customer resource management (CRM)
Video conferencing
IT service management
Accounting
Web analytics
Web content management

Benefits:

The biggest benefit of SaaS is costing less money than buying the whole application. The service provider generally offers cheaper and more reliable applications as compared to the organisation [1]. Some other benefits include Familiarity with the Internet, Better marketing, Smaller staff, reliability of the Internet, data Security, More bandwidth etc.

Obstacles:

SaaS isn't of any help when the organisation has a very specific computational need that doesn't match to the SaaS services

While making the contract with a new vendor, there may be a problem. Because the old vendor may charge the moving fee. Thus it will increase the unnecessary costs.

SaaS faces challenges from the availability of cheaper hardware's and open source applications.

2.2 Platform as a service (PaaS)

PaaS provides all the resources that are required for building applications and services completely from the Internet, without downloading or installing a software [1]. PaaS services are software design, development, testing, deployment, and hosting. Other services can be team collaboration, database integration, web service integration, data security, storage and versioning etc.

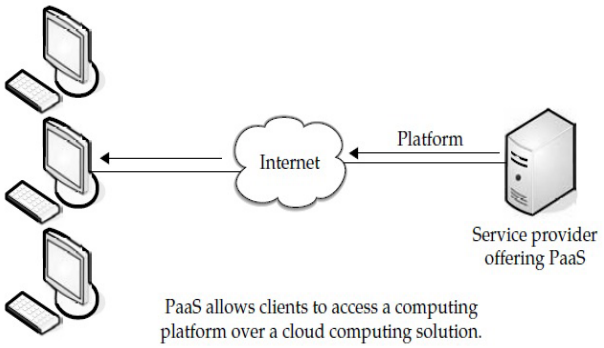

Fig 4: Platform as a service (PaaS)

Downfall :

lack of portability among different providers.

if the service provider is out of business, the user's applications, data will be lost.

2.3 Infrastructure as a service (IaaS)

IaaS, the capability provided to the consumer is to provision processing, storage, networks, and other fundamental computing resources where the consumer is able to deploy and run arbitrary software, which can include operating systems and applications. The consumer does not manage or control the underlying cloud infrastructure but has control over operating systems, storage, and deployed applications; and possibly limited control of select networking components (e.g., Host firewalls) [7] [5].

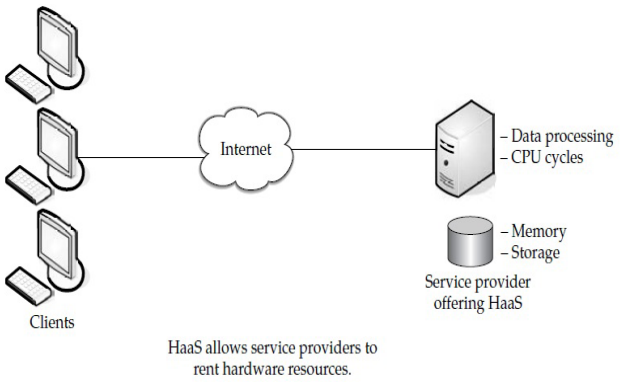

Fig 5: Infrastructure as a service (IaaS)

The increase in web traffic and different application in the web world is increasing day by day where millions of data are created every second, the Big Data is the big thing. Load balancing has become a very popular research field due to need of balancing the load on this heavy traffic [2]. Cloud computing as new computing technologies use virtual machine instead of physical machine to host, store and network the different nodes for their specific purpose. The load is done on CPU load, memory capacity, network. Load Balancing is done in such a way that all the load are distributed among various nodes in a distributive system. Failure of any node in the network will lead to unavailability of web resource in the web world. Load balancing should be able to provide scalability and availability [3]. Many authors agree with the definition of Cloud Computing as it consists of clusters of distributed computers (Clouds) providing on-demand resources or services over a network with the scale and reliability of a data centre [5] [6].

Load balancing:

It is a process of reassigning the total load to the individual nodes of the collective system to make resource utilization effective and to improve the response time of the job, simultaneously removing a condition in which some of the nodes are over loaded while some others are under loaded. A load balancing algorithm which is dynamic in nature does not consider the previous state or behavior of the system, that is, it depends on the present behavior of the system. The important things to consider while developing such algorithm are : estimation of load, comparison of load, stability of different system, performance of system, interaction between the nodes, nature of work to be transferred, selecting of nodes and many other ones [4] . This load considered can be in terms of CPU load, amount of memory used, delay or Network load.

Goals of Load balancing

The goals of load balancing are:

To improve the performance substantially

To have a backup plan in case the system fails even partially

To maintain the system stability

To accommodate future modification in the system

Types of Load balancing algorithms

Depending on who initiated the process, load balancing algorithms can be of three categories as: _ Sender Initiated: If the load balancing algorithm is initialized by the sender

Receiver Initiated: If the load balancing algorithm is initiated by the receiver

Symmetric: It is the combination of both sender initiated and receiver initiated

Depending on the current state of the system, load balancing algorithms can be divided into 2 categories as:

Static: It doesn't depend on the current state of the system. Prior knowledge of the system is needed

Dynamic: Decisions on load balancing are based on current state of the system. No prior knowledge is needed. So it is better than static approach.

Here we will discuss on various dynamic load balancing algorithms for the clouds of different sizes.

Dynamic Load balancing algorithm

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. The interaction among nodes to achieve load balancing can take two forms: cooperative and non-cooperative [4]. In the first one, the nodes work side-by-side to achieve a common objective, for example, to improve the overall response time, etc. In the second form, each node works independently toward a goal local to it, for example, to improve the response time of a local task. 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. A benefit, of this is that even if one or more nodes in the system fail, it will not cause the total load balancing process

to halt, it instead would effect the system performance to some extent. 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.

In non-distributed type, either one node or a group of nodes do the task of load balancing. Non-distributed dynamic load balancing algorithms can take two forms: centralized . In the first form, the load balancing algorithm is executed only by a single node in the whole system: the central node. This node is solely responsible for load balancing of the whole system. The other nodes interact only with the central node.

In semi-distributed form, nodes of the system are partitioned into clusters, where the load balancing in each cluster is of centralized form. A central node is elected in each cluster by appropriate election technique which takes care of load balancing within that cluster. Hence, the load balancing of the whole system is done via the central nodes of each cluster [4]. Centralized dynamic load balancing takes fewer messages to reach a decision, as the number of overall interactions in the system decreases drastically as compared to the semidistributed case. However, centralized algorithms can cause a bottleneck in the system at the central node and also the load balancing process is rendered useless once the central node crashes. Therefore, this algorithm is most suited for networks with small size.

Metrics in Load Balancing techniques in cloud computing:

*Throughput***:** It is used to calculate the no. of tasks whose execution has been completed. It should be high to improve the performance of the system.

Overhead: It determines the amount of overhead involved while implementing a load balancing algorithm. It is composed of overhead due to movement of tasks, inter-processor and inter-process communication. This should be minimized so that a load balancing technique can work efficiently.

Fault Tolerance: **It** is the time to migrate the jobs or resources from one node to other. It should be minimized in order to enhance the performance of the system.

Response Time: It is the amount of time taken to respond by a particular load balancing algorithm in a distributed system. This parameter should be minimized.

Resource Utilization: It is used to check the utilization of resources. It should be optimized for an efficient load balancing.

*Scalability***:** It is the ability of an algorithm to perform load balancing for a system with any finite number of nodes. This metric should be improved.

Performance: It is used to check the efficiency of the system. This has to be improved at a reasonable cost, e.g., reduce task response time while keeping acceptable delays.

III . ANALYSIS OF PROBLEM

The load balancer is able to handle half of the traffic in a cloud configuration to fully handle the traffic need to scale the numbers of the servers in the back end mainly for CPU-intensive applications [2]. To work efficiently and smoothly load balancing are implemented on different part of computing. Hardware load balancer and software load balancer are widely used, using of hardware load balancer is quite costly so some prefer to use software load balancer for their solution, open source load balancer which are widely used are Apache web Server's mod_proxy_balancer extension, Varnish, or the Pound reverse proxy and load balancer [3].

IV. PROPOSED WORK

In this seminar, Cloud computing is a distributed environment, by using semi-distributed load balancing method we can form the cluster and apply load balancing on the centre of the cluster. In this paper, it is noted about cloud computing and the purpose of load balancing in such an environment. Cloud computing uses the Virtualization technology. So now datacenter is equipped with lots of virtual machines. Applying the load balancing of virtual machine is not an easy task, by using my method of forming clusters of VMs we can easily do the load balance of Cloud Computing. Some basic algorithm for the entire process is applied from the existing literature. Clustering algorithm of PMs and VMs can be enhanced with the central node algorithm for the cluster. The objective of this paper is to design the concept of semi-distributed load balancing method for cloud computing. The main contribution of this paper is to use semi-distributed load balancing algorithm which can reduce the energy consumption in the data centre.

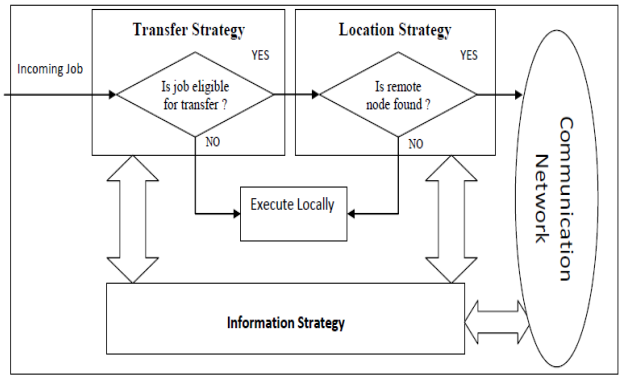

Fig 6. Interaction between components of a dynamic load balancing algorithm

Policies or Strategies in dynamic load balancing:

There are different policies in dynamic load balancing :

I. Transfer Policy: The part of the dynamic load balancing algorithm which selects a job for transferring from a local node to a remote node is referred to as Transfer policy or Transfer strategy.

*II. Selection Policy***:** It specifies the processors involved in the load exchange (processor matching)

*III. Location Policy***:** The part of the load balancing algorithm which selects a destination node for a transferred task is referred to as location policy or Location strategy.

*IV. Information Policy***:** The part of the dynamic load balancing algorithm responsible for collecting information about the nodes in the system is referred to as Information policy or Information strategy.

V. Load estimation policy: which determines how to estimate the workload of a particular node of the system.

*VI. Process transfer policy***:** which determines whether to execute a process locally or remotely.

*VII. Priority assignment policy***:** which determines the priority of execution of local and remote processes at a particular node.

*VIII. Migration limiting policy***:** which determines the total number of times a process, can migrate from one node to another.

V.SYSTEM MODELS

For the simulation, we assume the following system model. The cloud computing server is distributed across different geographical data centres, we assume a single data centre for our target system model. The target system model of *Mohsen et al.* [14] as a set named *P* consisting of *M* independent PMs that is presented as P = {*pm1, pm2,…, pmM*}. On each pmi, a Virtualization software (KVM, Xen) is installed that runs a set of *N* independent VMs represented by V ={*vm1,vm2,.., vmN*}. Ad centre architecture with a central node on VMs is presented in figure.7

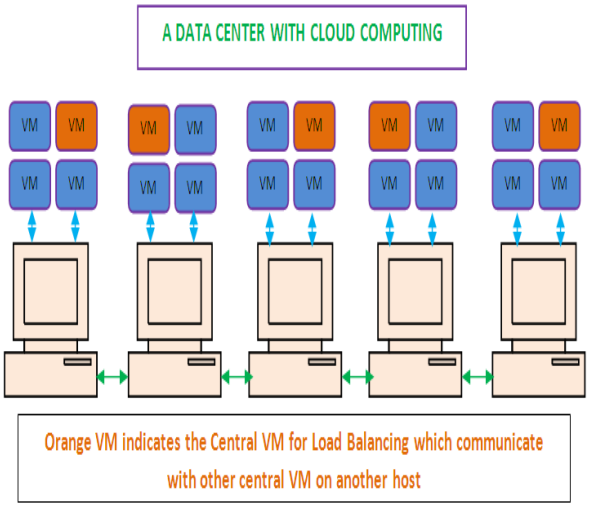

Fig 7. A Fully Interconnected Network Infrastructure

5.2 Algorithm Steps

Consider we have **P** as a Data Center

In a Data Center **m** number of the independent machine **PMs, as P ={*pm1, pm2,…, pmM*}**

Virtualization software on each **PMs**

Virtualization is installed that runs a set of **N** independent **VMs** represented by **V = {*vm1, vm2,.., vmN*}**

PMs are clustered with Hierarchical clustering to form a cluster

Central Node of **PMs** is elected with head Election by Fuzzy Logic for that cluster

VMs are clustered with Hierarchical clustering in the PMs cluster

Central Node of **VMs** is elected with head Election by Fuzzy Logic in **PMs** cluster

All tasks are queued at each central node which is VMs and are served on a First Come First Serve (FCFS) basis [4].

VMs Migration Process.

VI.APPLICATION

- **Google Drive**: Built off of Google Docs, Google Drive lets you port files among PCs, tablets and smartphones. Drive comes with 5 GB of free cloud storage. And, you can create docs, spreadsheets, presentations, drawings and more, just like Google Docs.
- **Microsoft SkyDrive**: This app allows cloud storage and file sharing through synced folders. You can view and edit SkyDrive files from any smartphone or tablet with the SkyDrive mobile app. Though it works for both Macs and PCs, additional services like settings backup and automatic photo upload to camera roll are available for Windows 8 and Windows Phone users.
- **DropBox:** DropBox allows you to sync files and folders across various platforms. The mobile app gives you access to all your DropBox files, which you can also share with other users. There's also a DropBox for teams that's designed for small businesses.

VII.IMPLICATION

In this paper, Cloud computing may be a distributed atmosphere, by victimization semi-distributed load equalization technique we will kind the cluster and apply load equalization on the centre of the cluster. during this paper, it's noted concerning cloud computing and also the purpose of load equalization in such associate atmosphere. Cloud computing uses the Virtualization technology. thus currently datacenter is provided with millions of virtual machines. Applying the load equalization of virtual machine isn't a simple task, by victimization my technique of forming clusters of VMs we will simply do the load balance of Cloud Computing. Some basic formula for the complete method is applied from the prevailing literature. clump formula of PMs and VMs may be increased with the central node formula for the cluster. the target of this paper is to style the construct of semi-distributed load equalization technique for cloud computing. the most contribution of this paper is to use semi-distributed load equalization formula which might cut back the energy consumption within the knowledge centre**.**

VIII.CONCLUSION

The purpose of this paper is to focus on one of the major concerns of cloud computing that is Load balancing. The goal of load balancing is to increase client satisfaction and maximize resource utilization and substantially increase the performance of the cloud system thereby reducing the energy consumed and the carbon emission rate. Also the purpose of load balancing is to make every processor or machine perform the same amount of work throughout which helps inincreasing the throughput, minimizing the response time and reducing the number of job rejection.

REFERENCES:

[1] Sandeep Tayal, "Tasks Scheduling optimization for the Cloud Computing Systems," International journal of advanced engineering sciences and technologies Vol No. 5, Issue no. 2, pp 111 – 115, 2011

[2] PabloValerio,"Load Balancing for Disaster Recovery," http:/ /content. dell.com /us /en / enterprise /large-business/load-balancing-disaster.aspx, 2011

[3] Load balancing (computing), "http://en.wikipedia.org/wiki/Load_balancing_(computing)"

[4] I.Ahmed and A.Ghafoor,"Semi-Distributed Load Balancing for Massively Parallel Multicomputers," IEEE Trans. Software Eng., Vol. 17, no. 10, pp 987-1004, 1991.

[5] Martin Randles, David Lamb, A. Taleb-Bendiab, "A Comparative Study into Distributed Load Balancing Algorithms for Cloud Computing," IEEE 24th International Conference on Advanced Information Networking and Applications Workshops, pp551-556, 2010

[6] Dilip A. Joseph, Arsalan Tavakoli and Ion Stoica, "A Policy-aware Switching Layer for Data Centers," SIGCOMM'08, pp51-62, 2008

[7] Ratan Mishra and Anant Jaiswal, "Ant colony Optimization: A Solution of Load balancing in Cloud," International Journal of Web & Semantic Technology Vol.3, No.2, pp33-50,2012

[8] Abbas Karimi, Faraneh Zarafshan, Adznan b. Jantan, A.R. Ramli1 and M. Iqbal b. Saripan, "A New Fuzzy Approach for Dynamic Load Balancing Algorithm," International Journal of Computer Science and Information Security, Vol. 6, No. 1, pp1-5, 2009

[9] Ali M. Alakeel, "A Guide to Dynamic Load Balancing in Distributed Computer Systems," International Journal of Computer Science and Network Security, Vol.10 No.6, pp153-160, 2010

[10] Network Load Balancing Technical overview, http://technet.microsoft.com/en- us/library/bb742455.aspx

[11] D. J. Evans and W.U.N. Butt," Dynamic load balancing using task transfer probabilities," Parallel Computing, Vol. 19, No. 8, pp. 897-916, 1993.

[12] S. Dhakal, M. M. Hayat, J. E. Pezoa, C. Yang, and D. Bader, "Dyanmic Load Balancing in Distributed System in the Presence of Delays: A Regeneration-Therory Approach,", IEEE Transactions on Parallel and Distributed Systems, vol. 18, no. 4, 2007.

[13] Z. Khan, R. Singh, J. Alam, and R. Kumar, "Performance Analysis of Dynamic Load Balancing Techniques for Parallel and Distributed Systems," International Journal of Computer and Network Security, vol. 2, no. 2, 2010

[14] Mohsen Sharifi, Hadi Salimi and Mahsa Najafzadeh, "Power-efficient distributed scheduling of virtual machines using workload-aware consolidation techniques," Springer Science+Business Media, LLC 2011

[15] Hierarchical clustering of networks,http: //en. wikipedia.org /wiki /Hierarchical _clustering _of_networks

[16] Indranil Gupta, Denis Riordan and Srinivas Sampalli, "Cluster-head Election using Fuzzy Logic for Wireless Sensor Networks", Proceeding CNSR '05 Proceedings of the 3rd Annual Communication Networks and Services Research Conference, pp255 - 260, 2005