An Approach on Semi-Distributed Load Balancing Algorithm for Cloud Computing System

Ajith Singh. N

Dept. of Computer Science
Karpagam University

M. Hemalatha

Dept. of Computer Science
Karpagam University

ABSTRACT

Cloud computing is deployed in the data centre where physical machine are virtualized. Cloud computing being the new technology has both advantages and disadvantages, one of the issues which cloud computing faces is load balancing. More than one virtual machine runs above the Virtualization. Load balancing in cloud computing is emerging topic which needs to be researched and study. The data centre is built with lots of systems where balancing is not an easy task especially for cloud computing. Most of the research is done in distributed environments. Using of semi-distributed load balancing in cloud computing is not discussed in any literature, wherever distributed load balancing on cloud computing is already in the list. By using the method of semi-distributed load balancing we can design a new algorithm for the cloud computing. This paper proposed to design a better load balance for the cloud computing which can be applied in every central node of the cluster.

Keywords

Cloud computing, Dynamic load balancing, Distributed systems, Algorithm, Semi-Distributed.

1. INTRODUCTION

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.

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 system 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
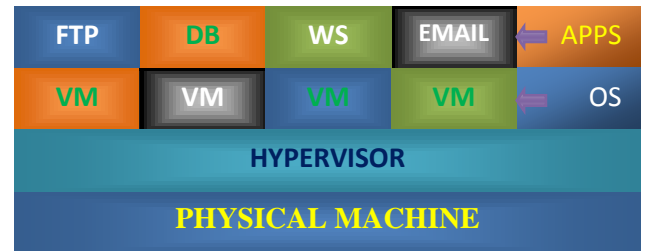

Figure 1. Virtualization on above the bare metal

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].

Other possible load balancing are used by DNS in which a single domain is mapped with different IP address where the client can choose which server to connect. The DNS method works well when the client is ready to locate their server on each continent and speed of data will be good because the client will be accessing the nearest server to it. The programmer is additionally applied load balancing method in their application for better flow of their application, the app is designed with many cookies, function and method to reduce the line of code and some are stored on a client computer for fast uploading, but if the client disable the cookies then there is no use of putting load balancing method of application [1]. When load balancer is implemented it should be in such way that the application needed. It can be said that based on the application different load balancer can be applied, it's not always correct that the for social network application round robin is best.

2. CLOUD COMPUTING

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)

SaaS, sometimes referred to as "on-demand software", is a software delivery model in which software and associated data are centrally hosted on the cloud. SaaS is typically accessed by users using a thin client via a web browser [7] [5].

2.2 Platform as a service (PaaS)

PaaS is a category of cloud computing services that provide a computing platform and a solution stack as a service. Along with SaaS and IaaS, it is a service model of cloud computing. In this model, the consumer creates the software using tools and/or libraries from the provider. The consumer also controls software deployment and configuration settings. The provider provides the networks, servers, storage and other services [7] [5].

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].

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].

3. LOAD BALANCING

Load Balancing are applied to computing resources (network, servers, hard drives) to solve a certain problem in a specific area. Different algorithm has been proposed by many researchers and has been discussed in many literatures [13] [4]. From the literature dynamic load balancing algorithm is applied either as distributed and non-distributed. The task of the load balancing is shared among the node in the distributed system.

Load balancing can take two form: cooperative and non cooperative. In the cooperative the nodes are working side-by-side to achieve a common objective were to improve the overall response time. In the non cooperative node works independently toward a goal like to improve the response time of a local task.

Benefit of dynamic load balancing is that if any node is fail it will not halt the whole network but it will affect the system performance. Node continuously interacts with each other which generate more message than non-distributed in the dynamic load balancing algorithms in distributed nature [4]. As the message is transmitted among the node for the interchange of system update it may lead to cause to system stress and may lead to affect system performance [11]. One node or a group of nodes makes the task of load balancing in non-distributed type, it can take two forms centralized and semi distributed.

In centralized form one node is exclusively responsible for the load balancing of the whole system that node is called as the central node. Central node is responsible to execute the load balancing algorithm [4]. In semi-distributed form a cluster is formed by a group of the nodes of the system, load balancing is happening in each cluster is of centralized form. Among the node in the cluster a central node initializes a to take care load balancing within that cluster. Semi distributed take more message as compared with centralized dynamic load balancing to reach a decision however centralized dynamic load balancing is more suited for network with small size which is not good for large network because load balancing is controlled by a central node if anything happen to the central node than whole network may fail or system may cause delay in processing the task [4] [7] [8].

Dynamic Load Balancing must have the four policies which are transfer policy, selection policy, location policy, information policy. Policy is responsible for keeping up-to-date load information about each node in the system. Information policy is used to maintain all the record of the node of the system. Transfer Policy has happened when a select job is needed for transfers from a local node to a remote node, Selection policy is involved while the processors involved in the load exchange, location policy is responsible to select the destination node for a transferred task is refereed to as location policy or location strategy [7] [8] [11]. Figure 2 demonstrates Interaction between components of a dynamic load balancing algorithm.

Semi-distributed policy acts as mediator between the centralized and the distributed policy, it's suitable for large distributed environment. Its produce better performance compared to centralized and distributed policies [4].

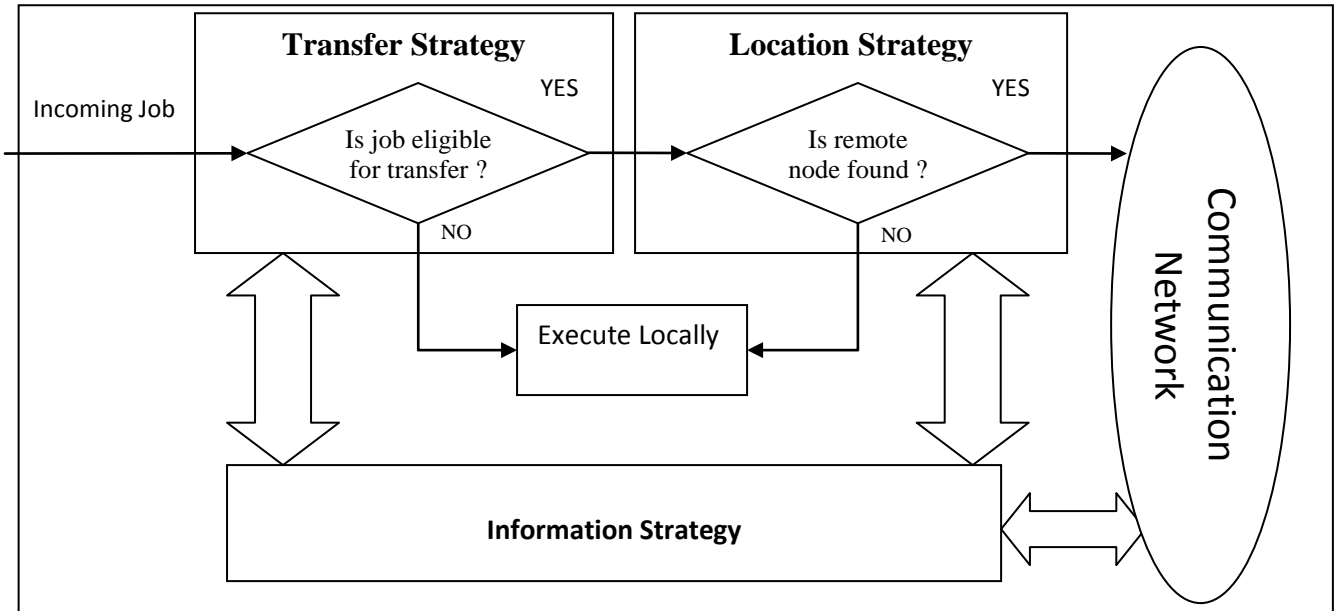The proposed work is based on semi-distributed policy.

Figure 2. Interaction between components of a dynamic load balancing algorithm (adapted from [9])

3. SYSTEM ASSUMPTIONS AND MODELS

3.1 Target System Model

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 = {pm_1, pm_2,…, pm_M}. On each pm_i, a Virtualization software (KVM, Xen) is installed that runs a set of *N* independent VMs represented by V ={vm_1,vm_2,.., vm_N}. Ad centre architecture with a central node on VMs is presented in Figure 3
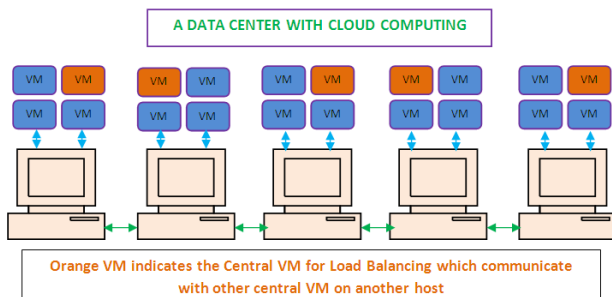

Figure 3. A Fully Interconnected Network Infrastructure Adopted from [14]

3.2 Assumptions

- PMs are all connected and are standalone machines which can be switched on or off independently.
- VMs can be paused, restart, or migrated to other PMs.
- PMs are clustered together by means of Hierarchical clustering [15] for the proposed model. Fig. 4 represent the hierarchical clustering dendrogram

- A central node is elected based on election policy cluster-head Election by Fuzzy Logic proposed by Indranil et al [16]
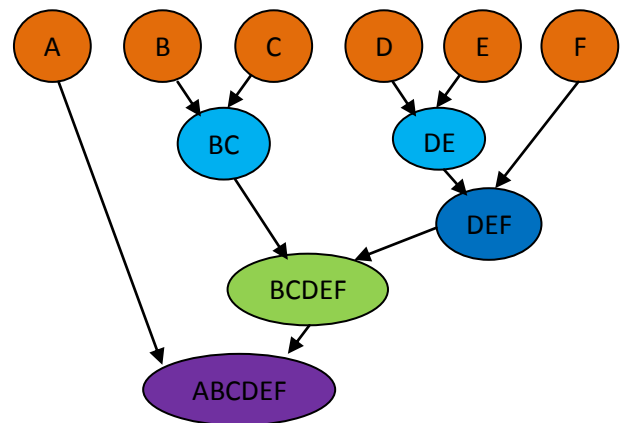

Figure 4. Hierarchical clustering dendrogram Adopted from [15]

3.3 Objective

When a Cluster head is selected for the central node to act as a load balancer in the semi-distributed system

- First its get all the load information of PMs and VMs
- Distance D is calculated among the host
- Load balancer in cloud system is applied
- Load Balancer continuously monitored of PMs and VMs.
- VMs migration is done from one PMs to another to save energy and cost.
- VMs migration only happens when PMs is processed only One VMs but if any PMs is not free in that cluster then migration does not happen.

3.4 Algorithm Steps

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

- In a Data Center **m** number of the independent machine **PMs**, as **P** ={$pm_1, pm_2,…, pm_M$}

- Virtualization software on each **PMs**

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

- **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

4. CONCLUSION

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.

5. FUTURE SCOPE

Future plan is to design clustering algorithm and load balancing for the cloud computing, using the simulation and other method. This gives a better result . Researchers can proceed and include security algorithm with load balancing.

6. ACKNOWLEDGMENT

7. 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] Pablo Valerio, "Load Balancing for Disaster Recovery," http://content.dell.com/us/en/enterprise/d/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-Theroy 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