Load Balancing Algorithm for Azure Virtualization with Specialized VM's

Anand Chaudhari
Department of Computer Science and Engineering
PRMIT&R, Badnera (Amravati), Maharashtra, India

Anushka Kapadia
Department of Computer Science and Engineering
PRMIT&R, Badnera (Amravati), Maharashtra, India

Abstract- One of the major problems that have been seen in computer networks and virtualization technique is of Network Congestion. Many techniques are available to overcome congestion problem but Time and space complexity is the issue. Time and space complexity increases when the Network Congestion takes place. The proposed work will provide a better solution to the azure OS load balancing by sorting request as per their type even if the specialized VM's crosses their threshold value the more created VM's are used by applying the Easily Spread Current Execution Algorithm. The best possible IJIET format paper on the above area of interest is tried and implemented which will sought some solution to the load balancing and distribution for azure OS. This paper presents an approach for scheduling algorithms that can maintain the load balancing and provides better improved strategies through efficient job scheduling and modified resource allocation techniques. As we have created multiple so that can have good back-up of virtual machines. After forwarding the request in first phase to their specialized VM's using round robin algorithm we will observe whether it crosses the threshold value of VM's and if it is crossing then the back-up VM's are used to handle the load using ESCE algorithm as shown below:

Keywords –Virtual Machine, Load Balancing, Easily Spread Current Execution

I. INTRODUCTION

Cloud computing is a fast growing area in computing research and industry today. It has the potential to make the new idea of computing as a utility' in the near future. The Internet is often represented as a cloud and the term "cloud computing" arises from that analogy. Cloud computing is the dynamic provisioning of IT capabilities (hardware, software, or services) from third parties over a network [3]. It is generally supposed that there are three basic types of cloud computing: Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS) [1]. In IaaS grids or clusters, virtualized servers, memory, networks, storage and systems software are delivered as a service. Perhaps the best known example is Amazon's Elastic Compute Cloud (EC2) and Simple Storage Service (S3), IaaS Provide access to computational resources, i.e. CPUs. And also provide (managed and scalable) resources as services to the user [3]. PaaS typically makes use of dedicated APIs to control the behavior of a server hosting engine which executes and replicates the execution according to user requests .E.g. Force.com, Google App Engine. Software as a Service (SaaS) Standard application software functionality is offered within a cloud. Examples: Google Docs, SAP Business by design. Load balancing is one of prerequisites to utilize the full resources of parallel and distributed systems. Load balancing mechanisms can be broadly categorized as centralized or decentralized, dynamic or static, and periodic or non-periodic. Physical resources can be split into a number of logical slices called Virtual Machines (VMs). All VM load balancing methods are designed to determine which Virtual Machine assigned to the next cloudlet [3]. This document introduce a new VMload balancing algorithm and compare the performance of this algorithms with the already existing algorithms like throttled and active monitoring VM load balancer [11]. Virtual machines (VMs) have the potential to play an important role in tomorrow's networked computing environments.

Load balancing is a relatively new technique that facilitates networks and resources by providing a maximum throughput with minimum response time [2]. Dividing the traffic between servers, data can be sent and received without major delay. Different kinds of algorithms are available that helps traffic loaded between available servers [1]. A basic example of load balancing in our daily life can be related to websites. Without load balancing, users could experience delays, timeouts and possible long system responses. Load balancing solutions usually apply redundant servers which help a better distribution of the communication traffic so that the website availability is conclusively settled .There are many different kinds of load balancing algorithms available, which can be

categorized mainly into two groups. The following section will discuss these two main categories of load balancing algorithms [1], [2].

Load Balancing in Cloud Computing:
Cloud vendors are based on automatic load balancing services, which allowed entities to increase the number of CPUs or memories for their resources to scale with the increased demands. This service is optional and depends on the entity's business needs. Therefore load balancers served two important needs, primarily to promote availability of cloud resources and secondarily to promote performance. According to the previous section Cloud computing will use the dynamic algorithm, which allows cloud entities to advertise their existence to presence servers and also provides a means of communication between interested parties [1].

II. CLOUD COMPUTING & VIRTUALIZATION

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

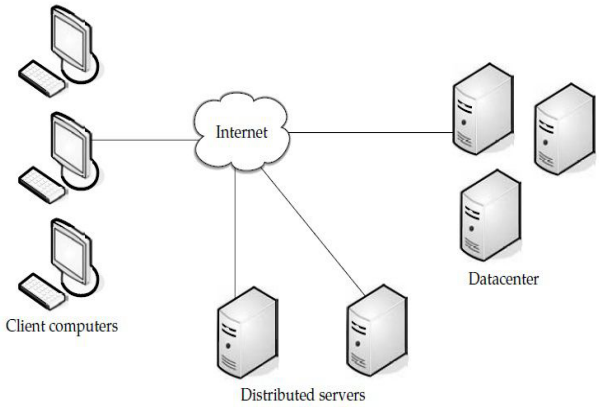

Figure 1. Fig 1: Three components make up a cloud computing solution [3]

A. Type of Clouds

1) Clients:
End users interact with the clients to manage information related to the cloud. Clients generally fall into three categories as given in [3]-[5]:

Mobile: Windows Mobile Smartphone, smart phones, like a Blackberry, or an iPhone.
Thin: They don't do any computation work. They only display 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.

2) 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 [5]. Now-a-days a concept called virtualization is used to install software that allows multiple instances of virtual server applications.

3) 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 [4].

Type of Clouds

Based on the domain or environment in which clouds are used, clouds can be divided into 3 catagories:
_ Public Clouds
_ Private Clouds
_ Hybrid Clouds (combination of both private and public clouds)
Virtualization
It is a very useful concept in context of cloud systems. Virtualization means "something which isn't real", but gives all the facilities of a real. It is the software implementation of a computer which will execute different programs like a real machine. Virtualization is related to cloud, because using virtualization an end user can use different services of a cloud. The remote datacenter will provide different services in a full or partial virtualized manner.

Two types of virtualization are found in case of clouds as given in [5]:
_ Full virtualization
_ Para virtualization

1) Full Virtualization:
In case of full virtualization a complete installation of one machine is done on another machine. It will result in a virtual machine which will have all the software's that are present in the actual server.

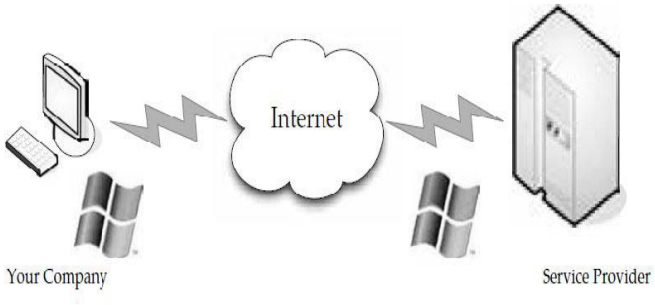

Figure 2. Full Virtualization [5]

Here the remote datacenter delivers the services in a fully virtualized manner. Full Virtualization has been successful for several purposes as pointed out in [5]:
- sharing a computer system among multiple users
- Isolating users from each other and from the control program
- Emulating hardware on another machine

2) Para virtualization:
In paravitualisation, the hardware allows multiple operating systems to run on single machine by efficient use of system resources such as memory and processor e.g. VMware software. Here all the services are not fully available, rather the services are provided partially [4].

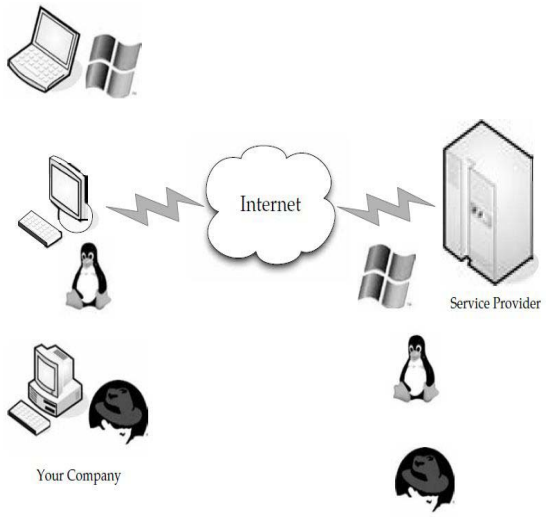

Figure 3. Para virtualization

Para virtualization has the following advantages as given in:

- Disaster recovery: In the event of a system failure, guest instances are moved to hardware until the machine is repaired or replaced.
- Migration: As the hardware can be replaced easily, hence migrating or moving the different parts of a new machine is faster and easier.
- Capacity management: In a virtualized environment, it is easier and faster to add more hard drive capacity and processing power. As the system parts or hardware's can be moved or replaced or repaired easily, capacity management is simple and easier.

III. PROPOSED WORK

Specialized Servers are created as VM's and they are arranged in a network will help for easy execution i.e. set-up including main server and cluster computers are made and then depending upon the request type the load is balanced.

As we have created multiple so that can have good back-up of virtual machines. After forwarding the request in first phase to their specialized VM's using round robin algorithm we will observe whether it crosses the threshold value of VM's and if it is crossing then the back-up VM's are used to handle the load using ESCE algorithm as shown below:

Note: Main Server is capable of handling all types of requests. Here we are considering the scenario that:

Request handling capacity of main server (load balancer) is full i.e. value crosses threshold limit of server and

Hereafter load balancing will be according to request type. Example Fax Request will be transferred to Fax server and so on.

For Example just consider that we have 4 types of requests as follows: Mail access Request, Media (Audio/Video) Request, Fax Request and Online-Chat Request. These users/clients request will be sorted and forwarded to specific server types without applying any load balancing algorithm. Hence the request handled by the clusters will be look:

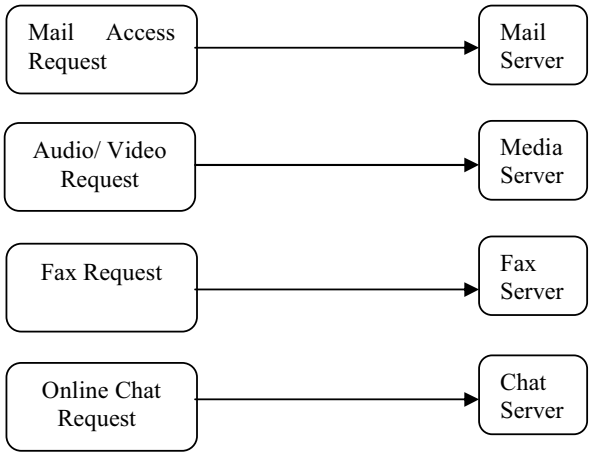

Figure 4. Load balancing using specialized server

As we have created multiple so that can have good back-up of virtual machines. After forwarding the request in first phase to their specialized VM's we will observe whether it crosses the threshold value of VM's and if it is crossing then the back-up VM's are used to handle the load using ESCE algorithm as shown below:

Round-Robin Algorithm

Round-robin is by far the simplest algorithm available to distribute load among nodes. It is therefore often the first choice when implementing a simple scheduler. On of the reasons for it being so simple is that the only information needed is a list of nodes [9], [15]. The algorithm traverses the list, returning the nodes one by one. When the end of the list is reached, the algorithm starts from the beginning of the list again [10].

Thus returning the same nodes in the same order. The time-complexity of the selection is O (1). Combined with its low implementation complexity and its low information requirements, the round-robin scheduler is also often the most efficient scheduler algorithm. However this is only when several key assumptions are true [3] – [7]:

1. The nodes must be identical in capacity. Otherwise performance will degrade to the speed of slowest node in the cluster.
2. Two or more client connections must not start at the same time. Should they, the node chosen will be the same, because the order of nodes retrieved from the cluster is the same every time.
3. The jobs must be similar to achieve optimum load distribution among the nodes. If a single node is more loaded than others it will become a bottleneck in the system.

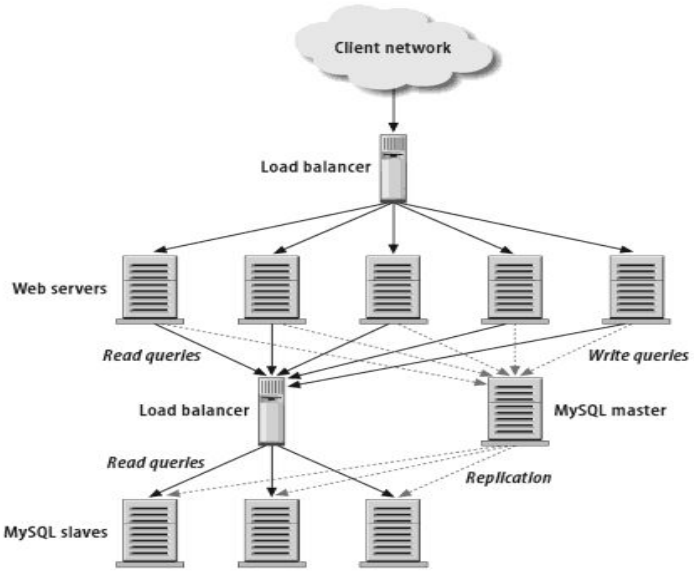

Figure 5. Load balancing using specialized server.

A. Equally Spread Current Execution

The random arrival of load in such an environment can cause some server to be heavily loaded while other server is idle or only lightly loaded. Equally load distributing improves performance by transferring load from heavily loaded server. Efficient scheduling and resource allocation is a critical characteristic of cloud computing based on which the performance of the system is estimated. The considered characteristics have an impact on cost optimization, which can be obtained by improved response time and processing time.

A scheduling algorithm is compared with the existing round robin scheduling to estimate response time, processing time, which is having an impact on cost .A Comparison of Dynamic Load Balancing Algorithms [6].

Here the jobs are submitted by the clients to the computing system. As the submitted jobs arrive to the cloud they are queued in the stack. The cloud manager estimates the job size and checks for the availability of the virtual machine and also the capacity of the virtual machine. Once the job size and the available resource (virtual machine) size match, the job scheduler immediately allocates the identified resource to the job in queue. Unlike the round robin scheduling algorithm, there is no overhead of fixing the time slots to schedule the jobs in a periodic way [6]. The impact of the ESCE algorithm is that there is an improvement in response time and the processing time. The jobs are equally spread, the complete computing system is load balanced and no virtual machines are underutilized. Due to this advantage, there is a reduce in the virtual machine cost and the data transfer cost [14] – [17].

ESAE LOAD ALGORITHM ACTIVE VM LOAD BALANCER [6]
[START] Step1:- find the next available VM
Step2:-check for all current allocation count is less than max length of VM list allocate the VM
Step3:- if available VM is not allocated create a new one in Step 4:- count the active load on each VM
Step5:- return the id of those VM which is having least load [END]

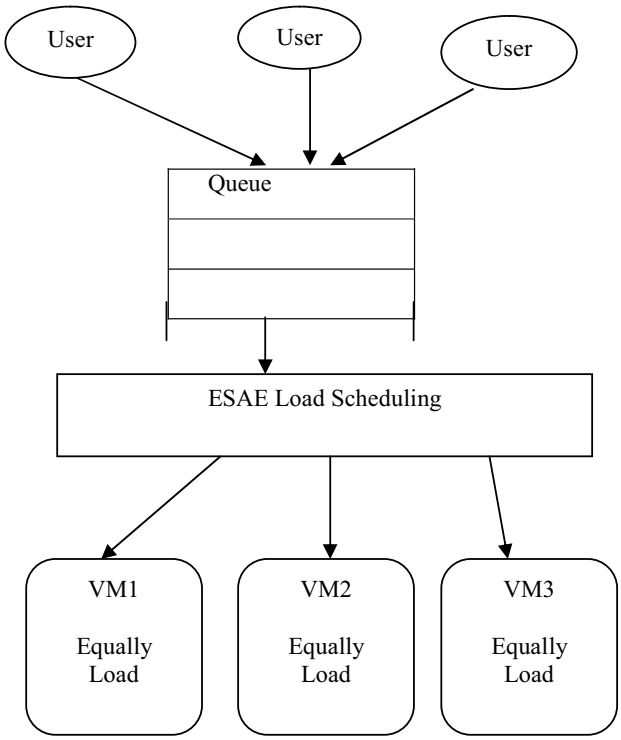

Figure 6. Figure 4: Equally spread Active execution load to the cloud system

IV.CONCLUSION

Load Balancing deals with cost and time challenge. Current strategies lack efficient scheduling and resource allocation techniques leading to increased operational cost and time. This paper aims towards the development of enhanced strategies through improved job scheduling and resource allocation techniques for overcoming the above-stated issues . In other case round robin algorithm is used for load balancing which proves to be an efficient and best algorithm as far as response time is concerned. Read/Write type client request are sorted and handled using model database. The server load balancing solution provides more advanced and flexible traffic management and stronger processing power as compared to single server implementation.

Here, Equal Spread Current Execution Load algorithm dynamically allocates the resources to the job in queue leading reduced cost in data transfer and virtual machine formation. In this paper we have proposed a new concept of azure OS load balancing with the help of specialized server and balancing request with the help of two algorithms. In first phase we are using round robin and the easily spread current execution. The paper describes how initial sorting of the request is done i.e. basically it is based on request type and then forwarded to specific type server. Combination of two algorithms provides better response time and handles cost and time complexities.

REFERENCES

[1] Zenon Chaczko, Venkatesh Mahadevan , Shahrzad Aslanzadeh Availability and Load Balancing in Cloud Computing University of Technology Sydney, Australia

[2] R. Shimonski. Windows 2000 & Windows Server 2003 Clustering and Load Balancing. Emeryville. McGraw-Hill

[3] B. Emin Gün Sirer, Robert Grimm, Arthur J. Gregory, Brian N. Bershad, “Design and implementation of a distributed virtual machine for networked computers”, University of Washington Department of Computer Science and Engineering.

[4] Meenakshi Sharma, Pankaj Sharma, Dr. Sandeep, “Sharma Efficient Load Balancing Algorithm in VMCloud Environment”, of CSE Sri Sai College of Engg. & Tech., Badhani, Pathankot, Punjab, India.

[5] Anthony T.Velte, Toby J.Velte, Robert Elsenpeter, Cloud Computing A Practical Approach, Tata McGraw-HILL Edition 2010.

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

[7] Ram Prasad Padhy, P Goutam Prasad Rao Load Balancing In Cloud Computing system Rourkela-769 008, Orissa, India May, 2011.

[8] David Karger and Matthias Ruhl, "New Algorithms for Load Balancing in Peer-to-Peer Systems," Tech. Rep. MIT-LCS-TR-911, MIT LCS, July 2000.

[9] Vinod Pisharody, Michael Wu, Tanya Zhiltsova Active-active operation for a cluster of SSL virtual private network (VPN) devices with load distribution.

[10] Anders Olaf Dannie, Michael palmeter Method and Apparatus For Distributing Load on Application Servers.

[11] [G. R. Andrews, D. P. Dobkin, and P. J. Downey, "Distributed allocation with pools of servers," in Proceedings of the first ACM SIGACT-SIGOPS symposium on Principles of distributed computing Ottawa, Canada: ACM, 1982, pp. 73-83.

[12] S. Malik "Dynamic Load Balancing in a Network of Workstation," 19 November 2000.

[13] N. G. Shivaratri, P. Krueger, and M. Singhal, "Load Distributing for Locally Distributed Systems," Computer, vol. 25, pp. 33-44, 1992.

[14] Rory Breuk, Gerrie Veerman Database Load Balancing, March 1, 2012.

[15] Tony Bourke, Server Load Balancing O,Reilly & Associates, Inc, August 2001.

[16] Andrew Tanenbaum, Computer Networks (Delhi, Pearson: Dorling Kindersley, 2008).

[17] Chandra Kopparapu, Load balancing Servers, Firewalls and Caches Wiley Computer Publishing 2002.