Efficient problem fixing for Virtual Machine Migration using Virtual Redundant Machine for Load Balancing in Cloud Computing

Rukman Palta[1], Rubal Jeet[2]

[1]Mtech (CSE), Indo Global College Of Engineering, Abhipur, Punjab Technical University, Jalandhar, India

[2]A.P (CSE), Indo Global College Of Engineering, Abhipur, Punjab Technical University, Jalandhar, India

Abstract— Cloud computing is a model that enables its end users a convenient and huge , on demand network access to multiple distributed computing resources, say, storage, network, CPU, disk drives, servers, applications, and services etc. which can be rapidly assigned and released with a minimal service provider interaction with end user. Load balancing is a mean for distributing workloads across multiple computing resources. Its main motive is to optimize the usage of resources, maximizing Throughput, minimizing Response Time, and also avoiding unnecessary overload of any such resources. It becomes a serious problem with the continuous increase in the number of users and type of applications on clouds. This paper presents a new design which results into better resource management of Virtual machine and also results into balancing the load of data centers. With this method even with infinite user requests, the data centers will continue working and thus there is no chance of their failure.

Keywords- *Cloud Computing; Load balancing; Virtualization; Virtual Machine Migration.*

I. INTRODUCTION

Cloud computing is a combination of - *Software as a service* (SaaS), *Platform as a service* (PaaS) and Infrastructure *as a service* (IaaS). It is also named as *XaaS* that means *anything as a service*. As *IaaS*, it provides hardware and network facility to the end user; thus end user will itself installs or develops its own OS, Software and application. As *SaaS*, cloud offers a pre-made application along with required software, OS, Hardware and network facility. As *PaaS*, end users are given OS, Hardware, and Network facilities and it is the end user which installs or develops its own software and application.

II. OVERVIEW OF LOAD BALANCING AND PROBLEM FORMULATION

The goal of load balancing is to improve the performance by balancing the load among various resources such as network links, central processing units, disk drive to achieve optimal resource utilization, maximizing throughput, maximizing response time, and avoiding overload and under load *situations*. Load balancing is necessary to share the dynamic load across numerous nodes to make sure that no single node is overloaded or idle. Many techniques and algorithms had been established in past but some were having problem with throughput, some resulted into failure of data centers in case of large user requests. Virtual Machine Migration plays vital role in attaining load balancing in distributed environment such as clouds. Existing algorithms do not always give expected performance with large scale and service-oriented data centers. VMM provides solution for load balancing in various deployment model, so there arise a need to overcome this challenge. Better response time, minimized downtime and less workload is utmost requirement in Virtual Machine Migration.

A. Related Work

[3] Provided the advantages and disadvantages of various existing load balancing algorithms by comparing the existing algorithms such as CLDBM(*Central Load Balancing Decision Model*), ANT COLONY, MapReduce, WLC (*weighted least connection*), DDFTP(*dual direction downloading algorithm from FTP servers*) and VM mapping . It also discussed the challenges that must be addressed to provide most suitable and efficient load balancing algorithms.**[5]** Investigated various techniques developed for migrating a whole virtual cluster. This paper also presented several challenges in live migration of virtual clusters such as huge amount of data, Limitation of network bandwidth, Communications between VMs. Experimental results reveal —

1. Average migration time was sensitive to the memory size, that is, with increase in the memory size there was considerable increase in the total migration time.
2. Downtime, however, remained unaffected with the change in memory size as downtime is affected by the rate of dirty pages and transferred pages.
3. Both total migration time and downtime increased linearly as the number of migrations nodes increases.

4. Virtual clusters showed good scalability as both average migration time and average downtime decrease as the number of migration nodes increases.
5. In case when Concurrent migration granularity increased to 16, HPL throughput decreased dramatically as 16 virtual machine nodes were migrated simultaneously, this in turn resulted into overloaded network bandwidth.
6. In case of Mutual migration, sequential migration was better than Concurrent migration. Also Unidirection migration performance was better than the bidirection migration performance due to limited network bandwidth.
7. Keeping the concurrent granularity small, bidirection migration performance was very close to that of the unidirection one. But with the increase in concurrent granularity, the performance of bidirection migration decrease quickly.
8. In case of Live migration of homogeneous Virtual Clusters, sequential migration performed better then the concurrent one. Also *Node by node* migration was slightly better than *Cluster by cluster* migration due to less synchronization overheads between each virtual machine in each virtual cluster.

III. VIRTUALIZATION

Virtualization is a mechanism under which the resources of a computer are divided into multiple execution environments. Virtualization is an illusion which actually gives the services of a reality. *Migration* of a virtual machine is moving or replicating the jobs running on a physical machine (source node) to another physical machine (target node). While the VM is running on the source node, without disrupting any active network connections, or disrupting the normal system activities, the VM is moved to the target node. The main reasons of Virtual Machine Migration are:

a) Load Balancing
b) Maintanence

IV. RESEARCH OBJECTIVES

1. Implement VRM (virtual Redundant MACHINE) that could come into the existence when physically machine is heavily loaded and provide better resource utilization of Vritual machines.
2. Simulate the Process in NS2.
3. Compare the results with earlier technique and analyzing results using VRM.

V. PROPOSED WORK

A good system is one which results into minimum Response time, minimum Migration time and less workload due to migration process. For this a system is designed which is composed of several replicated machines called "**Virtual redundant machine**. Load would be migrated automatically to the attached VRMs once the data centers in a particular zone get overloaded. This will fulfill our purpose of attaining Virtual machine migration which is one of the major challenges for load balancing mechanism. It will cause Better Process Management, better Resource allocation and better response time to end users. Even infinite user requests will affect VRMs only. Chances would be rare that a data center could fail even with infinite user requests.

A. Basic Design of the Proposed System

The basic design of the system is shown in figure 5.1. Here considering four zones where each zone contains, say, three data centers. With each zone there is also an attached **VRM (Virtual redundancy machine)**. Initially all the data centres are under loaded. Slowly load at each data centre is increased as user requests for various processes increase. This process continues until it reaches a point where all the data centres at a zone get overloaded and reaches a threshold value. After reaching a threshold limit, the load from these data centres is balanced by migrating this load onto the attached VRM. All the processes that are running on the data centres are copied onto the VRM. The source machine continues running while certain pages are pushed to the new destination VRM. To ensure consistency, pages modified during this process will be re-sent. Thus the overloaded physical machine returns to its normal state by pushing some of its load to VRM.

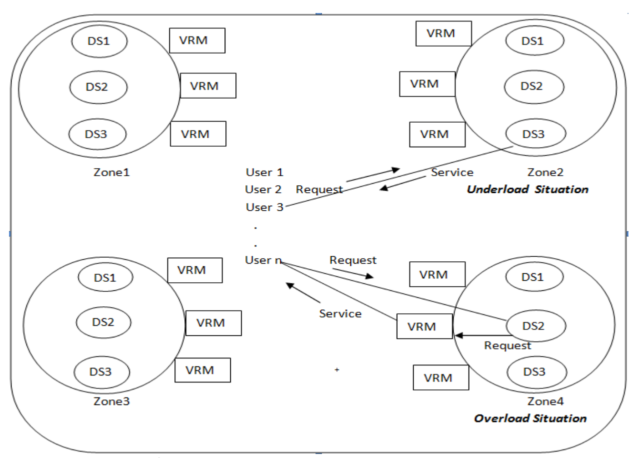

Fig.5.1: Design of Virtual Redundancy Machine

B. Stack Partitioning Structure

Main idea behind this research work is its stack partitioning structure. Whenever a particular data centre gets overloaded, it will send request to first attached VRM, if it is free then it will provide service to the end user as if data center itself is providing service. During next iteration on a request from data centre load is pushed to second attached

VRM not to first. Thus after three iterations all the three VRMs will be serving data centres, not just one. On next user request, again load is pushed to first VRM by partitioning its stack into two equal parts. At fifth iteration stack of 2nd VRM is partitioned into two equal parts and similarly at 6th iteration stack at 3rd VRM will b partitioned into two equal parts as well. On next round at 7th iteration, stack at 1st VRM will b partitioned into three equal parts. This process continues untill all the data centres at that particular zone will not get underloaded. The significance of this design is that it will not only balance the load of data centres, but it will also, manage to utilize the resources and availability of VRMs. Chances of VRM failure would automatically reduce as all the three VRMs will help balancing the load from Data Centres. Single VRM is not responsible for balancing load from data centres. All the VRMs will unload the data centres by partitioning its stack. Thus load is first transfer to 1st VRM then to 2nd VRM and then to 3rd VRM and finally again to 1st VRM. This will also allow VRMs to unload their load while others are busy serving the end users via data centres and if data center is free from all the requests then, loaded vrms will pop the request and send it to the data center.

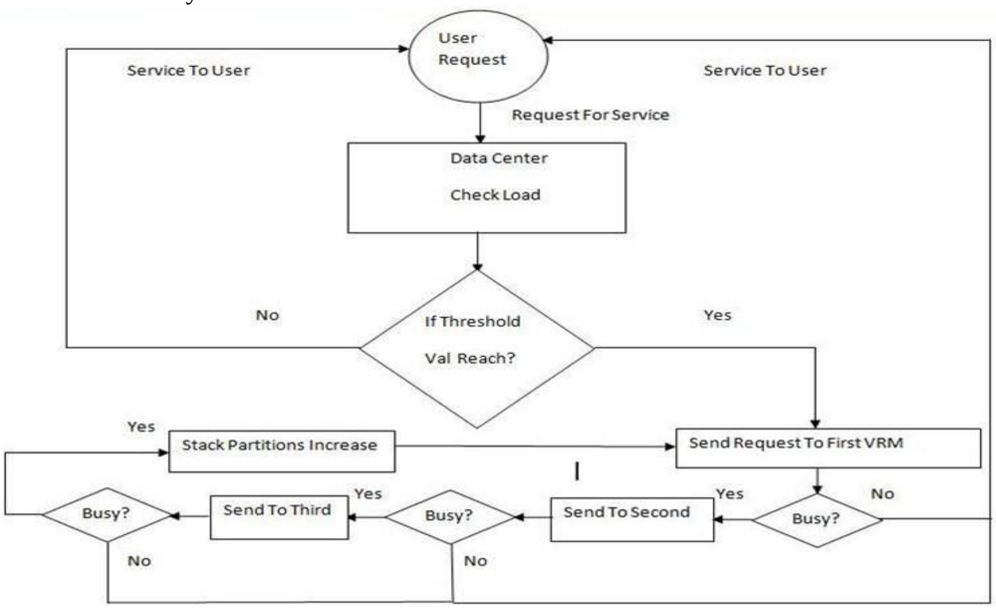

Figure 5.2: Flow Chart of Load Balancing Using Virtual Machine Migration

VI. RESULT AND DISCUSSION

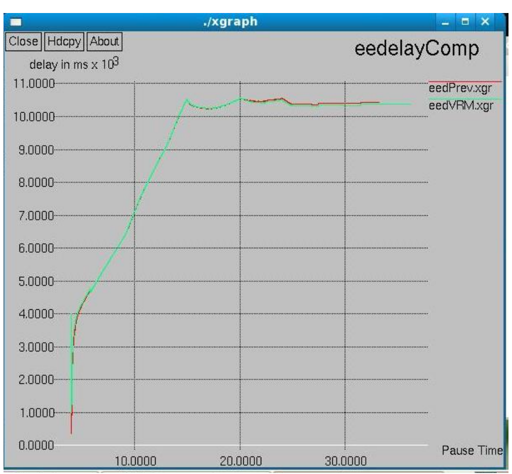

Fig1: eedelayPrev>eedelayVRM

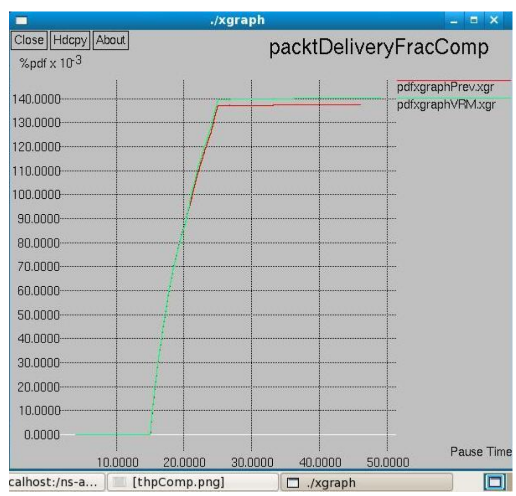

Fig2: pdfPrev<pdfVRM

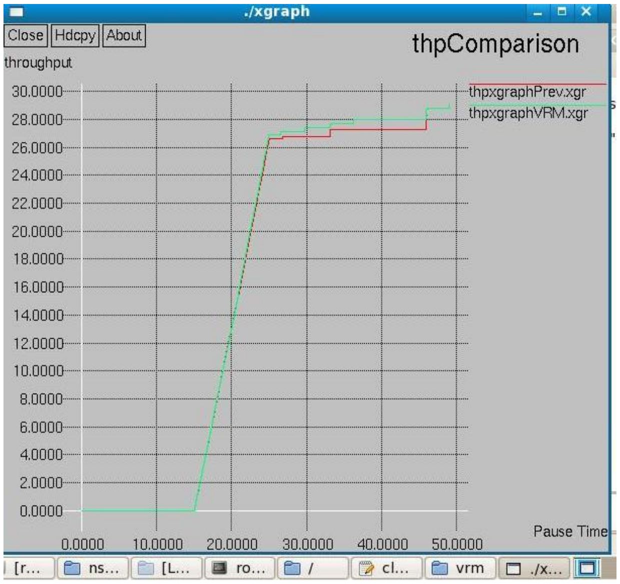

Fig3:thpPrev<thpVRM

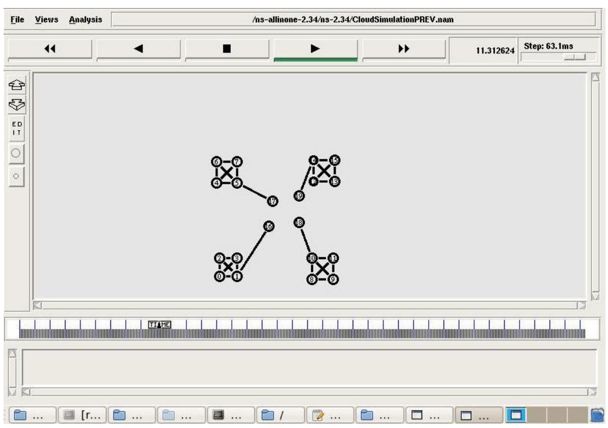

Fig4: WithoutVRMSimulation

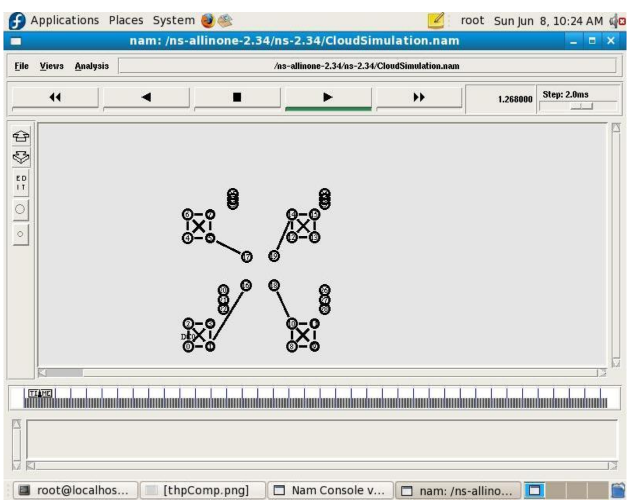

Fig 5: WithVRMSimulation

VII. CONCLUSION AND FUTURE SCOPE

This research presents a new structure for balancing load in Cloud environment using Virtual Redundant Machine that is based on the concept of virtual machine migration. With this structure the end-to-end delay is measured to be less in case of system with attached VRMs as compared to system without VRM. Similarly the average throughput and packet delivery fraction is higher in case of system with VRM as compare to system without VRM. The obvious reason behind this performance is the structure employed in this work. In earlier system either the whole virtual cluster was migrated which causes large migration time and lesser throughput or else system with one VM, where, a single VM was responsible for balancing load of the system. This could result into VM failure due to large user requests. But here each zone incorporates three VRMs and at a time the load is being shared n balanced via all the three attached VRMs by stack partitioning structure. This cause not only balancing load in the Data centres but also helps in utilizing the resources of all the VRMs and availability of all the VRMs. Here this system support only three VRMs in a particular zone with each having a stack partitioning upto 12. In future we can enhance it to support n user requests which would further raise the system performance.

VIII. REFERENCES

[1] A.Khiyaita, M.Zbakh, H. El Bakkali, Dafir El Keytani "Load Balancing Cloud Computing : State of Art" Proceedings of IEEE 2012.

[2] Haozheng Ren, Yihua Lan, Chao Yin "The Load Balancing Algorithm in Cloud Computing Environment" Proceedings of 2012 IEEE and 2012 2nd International Conference on Computer Science and Network Technology.

[3] Klaithem Al Nuaimi, Nader Mohamed, Mariam Al Nuaimi and Jameela Al-Jaroodi "A Survey of Load Balancing in Cloud Computing: Challenges and Algorithms" 2012 IEEE Second Symposium on Network Cloud Computing and Applications.

[4] Kejiang Ye et. al. "Live Migration of Multiple Virtual Machines with Resource Reservation in Cloud Computing Environments" 2011 IEEE 4th International Conference on Cloud Computing.

[5] Kejiang Ye, Xiaohong Jiang, Ran Ma, Fengxi Yan "VC-Migration: Live Migration of Virtual Clusters in the Cloud" 2012 ACM/IEEE 13th International Conference on Grid Computing.

[6] Divya Kapil, Emmanuel S. Pilli and Ramesh C. Joshi "Live Virtual Machine Migration Techniques: Survey and Research Challenges" Proceedings of 2012 IEEE and 2013 3rd IEEE International Advance Computing Conference (IACC).

[7] Raghavendra Achar et. al. "Load Balancing in Cloud Based on Live Migration of Virtual Machines" 2013 Annual IEEE India Conference (INDICON).

[8] 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.

[9] Sachin Kumar, Niraj Singhal "A Priority based Dynamic Load Balancing Approach in a Grid based Distributed Computing Network" International Journal of Computer Applications, July 2012.
[10] Yi Zhao "Adaptive Distributed Load Balancing Algorithm based on Live Migration of Virtual Machines in Cloud" 2009 Fifth International on INC.

Rukman Palta is currently doing her M.tech in Computer science and engineering from Indo Global College of engineering, Abhipur Mohali. She had done her B.tech in Computer science in 2011. Her areas of interest are Grid Computing, Cloud Computing, Virtualization, Algorithms, Data structure.