ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 2, February 2014

Load Balancing and Maintaining the Qos on Cloud Partitioning For the Public Cloud

[1]S.Karthika, [2]T.Lavanya, [3]G.Gokila,[4]A.Arunraja [5]S.Sarumathi, [6]S.Saravanakumar, [7]A.Gokilavani

[1,2,3,4]Student, Department of Computer Science and Engineering, Jay Shriram Group of Instutions, Avinashipalayma, Tirupur, Tamilnadu, India

[5,6, 7]Assistant Professor, Department of Computer Science and Engineering, Jay Shriram Group of Instutions, Avinashipalayma, Tirupur, Tamilnadu, India

ABSTRACT : The load balancing in the cloud architecture is a important performance for the cloud environment. The advanced and prior load balancing will provided tremendous advantages to the cloud users. In case of a client server architecture still the architecture need some improvements. The improvements basically need the query based wireless sensor networks. The main objective of our project is to develop a query-based wireless sensor systems, where a user would issue a query and expect a response to be returned within the deadline.In addition to this, we also develop an adaptive fault-tolerant quality of service control methods which is based on hop-by-hop data delivery utilizing "source" and "path" redundancy, with the goal to satisfy application of QoS requirements which prolongs the lifetime of the sensor system . This project explains about the better architecture of the cloud switching in different mechanism and also effectively balances the load with Q OS and query aggregation process.

KEYWORDS*:*Adaptive fault tolerance ; Query aggregation; Load balancing algorithm; Wireless sensor system

I. INTRODUCTION

Cloud computing is a new paradigm in which computing resources such as processing, memory, and storage are not physically present at the user's location. Instead, a service provider owns and manages these resources, and user accesses them via the Internet. Cloud computing is a general term for anything that involves delivering hosted services over the Internet. These services are broadly divided into three categories: Infrastructure-as-a-Service (IaaS), Platform-as-a- Service (PaaS) and Software-as-a-Service (SaaS) A cloud can be private or public. A public cloud sells services to anyone on the Internet. A private cloud is a proprietary network or a data center that supplies hosted services to a limited number of people. When a service provider uses public cloud resources to create their private cloud, the result is called a virtual private cloud. Private or public, the goal of cloud computing is to provide easy, scalable access to computing resources and IT services.For example, Amazon Web Services lets users store personal data via its Simple Storage Service (S3) and perform computations on stored datausing the Elastic Compute Cloud (EC2). This type of computing provides many advantages for businesses—including low initial capital investment, shorter start-up time for new services, lower maintenance and operation costs, higher utilization through virtualization, and easier disaster recovery—that make cloud computing an attractive option. Reports suggest that there are several benefits in shifting computing from the desktop to the cloud.

Load balancing 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 behaviour of the system, that is, it depends on the present behaviour of the system. The important things to consider while developing such algorithms 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.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

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 2, February 2014

- To maintain the system stability
- To accommodate future modification in the system

II.RELATED WORK

There have been many studies of load balancing for the cloud environment. Load Balancing Model Based on Cloud Partitioning for the Public Cloud [1]concept with a switch mechanism to choose different strategies for different situations. The algorithm applies the game theory to the loadbalancing strategy to improve the efficiency in the public cloud environment.The NIST definition [2] characterizes important aspects of cloud computing and provide a baseline cloud computing.Computer architectures [3] should shift the focus ofMoore's law from increasing clock speed per chip to increasing the number of processor cores and threads per chip. Enhanced Load Balancing Approach to Avoid Deadlocks in Cloud [4] avoid deadlocks among the Virtual Machines (VMs) while processing the requests received from the users by VM migration. The deadlock avoidance enhances the number of jobs to be serviced by cloud service provider and thereby improving working performance and the business of the cloud service provider.Checkpoint-based Intelligent Fault tolerance For Cloud Service Providers [5] proposes a smart checkpoint infrastructure for virtualized service providers and fault tolerance model for real time cloud computing. The checkpoints are stored in a Hadoop Distributed File System. One advantage of cloud computing is the dynamicity of re- source provisioning.Analysis of Fault Tolerance Approaches [6] in Dynamic Cloud Computing proposed a method for dynamic load balancing technique which is used to avoid this fault tolerance in cloud computing. The Dynamic Load Balancing algorithm checks the utilization of the CPU, if CPU has less utilization given in the algorithm can response the client request otherwise the request is shift to another server with the help of load balancer. This technique gives the better result. It deals to analysis the fault tolerance method by using fault tolerance algorithm.A QOS Control Method Cooperating with a Dynamic Load Balancing Mechanism[7] describes about the dynamic traffic engineering architecture and provides QoS-guaranteed service (GS), in addition to existing best effort (BE) service. It is designed to provide bandwidth guaranteed paths for GS traffic along optimum routes and to provide one or more paths for BE service traffic along routes that will best avoid congestion.. Use of this architecture will help to provide QoS-guaranteed service, effectively utilize network resources, and avoid degradation in BE traffic throughput.

III. SYSTEM MODEL

There are several cloud computing categories with this work focused on a public cloud. A public cloud is based on the standard cloud computing model, with service provided by a service provider. A large public cloud will include many nodes and the nodes in different geographical locations. Cloud partitioning is used to manage this large cloud. A cloud partition is a subarea of the public cloud with divisions based on the geographic locations. The architecture is shown in Fig.1The dotted line around the Blade server/Top of rack (ToR) switch indicates an optional layer, depending on whether the interconnect modules replace the ToR or add a tier.

The load balancing strategy is based on the cloud partitioning concept. After creating the cloud partitions, the load balancing then starts: when a job arrives at Fig. 1Typical cloud partitioning. The system, with the main controller deciding which cloud partition should receive the job. The partition load balancer then decides how to assign the jobs to the nodes. When the load status of a cloud partition is normal, this partitioning can be accomplished locally. If the cloud partition load status is not normal, this job should be transferred to another partition as server virtualization and client virtualization.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 2, February 2014

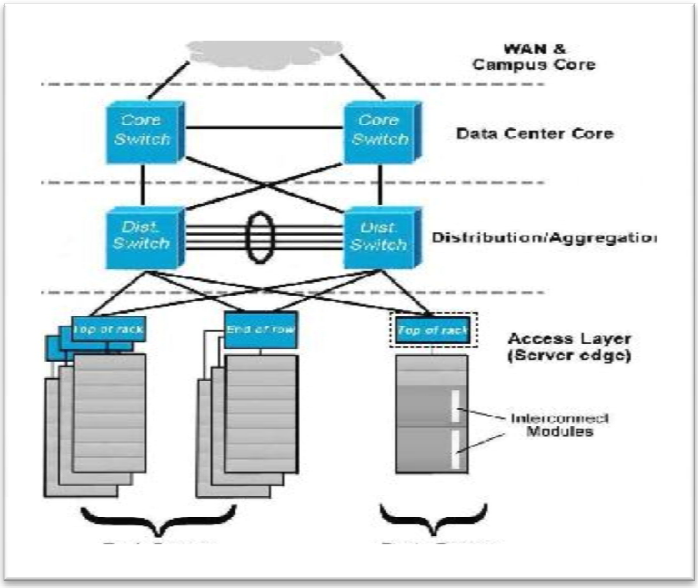

Fig.1 Typical Data Centre Structure with three layers

A. SERVER VIRTUALIZATION

The increase in VM density and the significant increase in VM mobility have resulted in greater performance demands on the network subsystems at the server-network edge. Moving workloads dynamically requires VMs to stay within a common VLAN in the same Layer 2 (L2) network. If you want to move a VM outside its L2 domain, you have to use manual processes such as assigning and updating the I P addresses for the failed-over services and updating DNS entries correctly. To provide maximum VM flexibility, many enterprises are evaluating ways to enlarge their L2 networks.

B. CLIENT VIRTUALIZATION

A specialized type of VM is the client virtualization technology such as virtual desktop infrastructure (VDI). VDI creates a client desktop as a VM. It includes the real-time compilation of the end user's data, personal settings, and application settings with a core OS instance and a shared generic profile. You can either install the end-user applications locally as a fully packaged instance or stream them from outside the VM. It improves performance and reliability by using mostly cable-free internal chassis connections between hosts and management services.

C. PRACTICAL SOLUTIONS FOR OPTIMIZING TRAFIC FLOW

- Fostering E/W traffic flow at the physical server-network edge .
- Distributing management intelligence at the physical server-network edge rather than concentrating it at higher layers in your network.
- Flattening your L2 network by using technologies like HP Intelligent Resilient Framework™ (IRF).
- Making your L2 network more efficient by implementing future multi-path standards.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 2, February 2014

D. IDENTIFY TRAFIC BOTTLENECK

- Tradeoffs can be made depending on two criteria:
- Whether we want to optimize the E/W traffic flow by providing intelligent management at the physical switch-server edge
- Whether we want to optimize for performance inside a physical server between multiple VMs (with possible degradation of network management visibility and control)

E. VIRTUAL SWITCH ARCHITECTURES

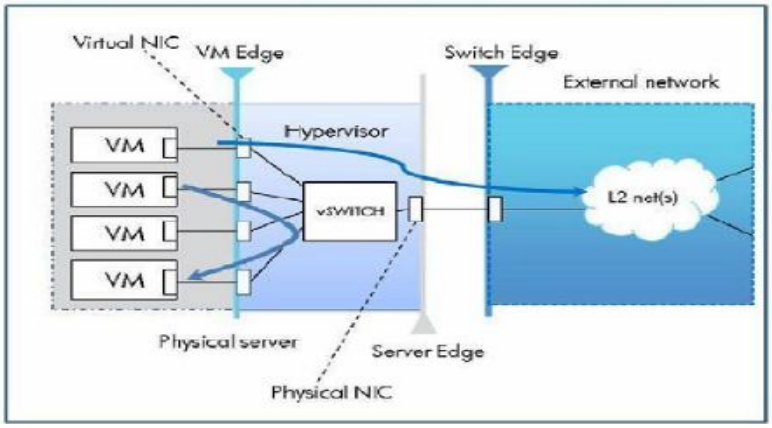

Fig.2. VSwitches internal VM -VM traffic

However, there are some limitations to a vSwitch:

- It moves the control point for networking infrastructure into the domain of the server administrator. This management stack is typically a component of the server-based hypervisor tool aimed at system and virtualization administrators. As such, vSwitch management generally does not integrate with existing external physical network policy and management tools. This usually means two different teams (with different processes) manage the physical network and the virtual network, even though the management tasks and functionality overlap.

- It consumes valuable CPU bandwidth. The higher the traffic load, the greater the number of CPU cycles required to move traffic through the vSwitch. This reduces the ability to support larger numbers of V M s in a physical server.

- It lacks network-based visibility. A vSwitch does not have standard network monitoring capabilities such as flow analysis, advanced statistics, and remote diagnostics of external network switches. When network outages or problems occur, identifying the root cause can be difficult in a virtualized server environment.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 2, February 2014

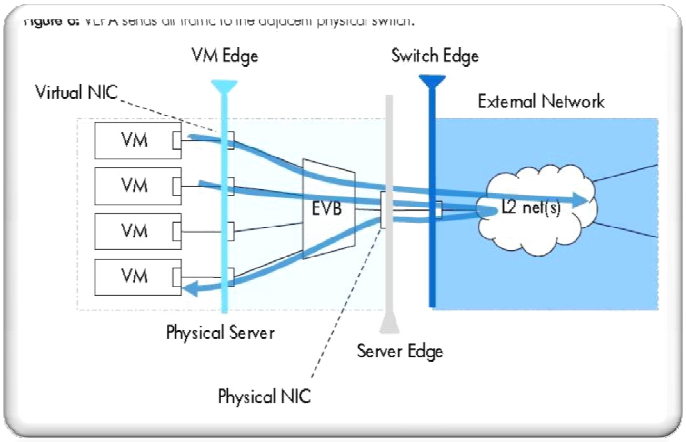

Fig.3.VEPA sends all traffic to the adjacent physical switch.

Advantages of VEPA include:

- Moves the VM connection control point into the edge physical switch (ToR or EoR). VEPA leverages existing investments made in data center edge switching. Administrators can manage the edge network traffic using existing network security policies and tools.
- Offloads the server's CPU from the overhead related to virtualization specific network processing and forwarding
- Improves security. Most ToR switches support hardware-based access control lists (TCAM s), allowing thousands of these filters to be processed without any effect on performance.
- Improves visibility. Monitoring technologies likesFlow in the edge switch can provide a full, end-toend understanding of traffic flows.

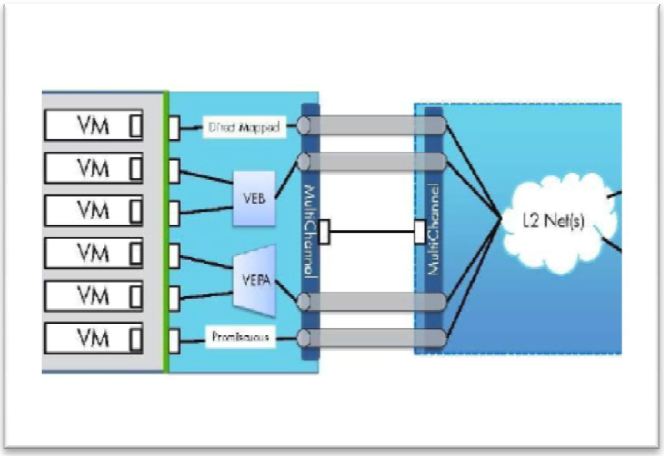

Fig 4:VEPA Multi channel Capabilities

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 2, February 2014

F. HP VIRTUAL CONNECT

One of the ways to optimize the server edge for E/W traffic flow is by implementing H P Virtual Connect technology. Virtual Connect is a set of interconnect modules and embedded software for HP BladeSystem c- Class enclosures that provides server-edge and I/O virtualization. It delivers direct server-to-server connectivity within an enclosure² especially important for the latency sensitive, bandwidth-intensive applications that we've been discussing. For example, as described in the Client Virtualization section, using BladeSystem with Virtual Connect lets you design an infrastructure that can optimize network traffic without leaving the enclosure.

IV. CONCLUSION

Thus the Survey to develop a query-based wireless sensor systems, where a user would issue a query and expect a response to be returned within the deadline is studied.

In addition to this, we also develop an adaptive fault-tolerant quality of service control methods which is based on hop-by-hop data delivery utilizing "source" and "path" redundancy, with the goal to satisfy application of QoS requirements which prolongs the lifetime of the sensor system is explained. Thus our project also explains the better architecture of the cloud switching in different mechanism and also effectively balances the load with QOS and query aggregation process.

REFERENCES

1. GaochaoXu, Junjie Pang, And XiaodongFu,A Load Balancing Model Based On Cloud Partitioning For The Public Cloud, IEEE Transactions On Cloud Computing ,Issn 1007-0214 04/12 pp34-39 ,Volume 18, Number 1, February 2013

2. Peter Mell,Timothy Grance,The NIST Definition Of Cloud Computing,Special Publication 800-Recommendations Of The National Institute Of Standards And Technology,Vol. 25, No. 12, Pp. 33-44,September 2011.

3. S. Penmatsa And A. T. Chronopoulos, Game-Theoretic Static Load Balancing For Distributed Systems, Journal Of Parallel And Distributed Computing, Vol. 71, No. 4,Pp. 537-555, Apr. 2012.

4. Rashmi. K S, Suma. V And Vaidehi.M., Enhanced Load Balancing Approach To Avoid Deadlocks In Cloud ,IJCA Journal ACCTHPCA, Vol2, Number2,Pp 200-213 ,Year Of Publication2012 .

5. N. G. Shivaratri, P. Krueger, And M. Singhal, Load Distributing For Locally Distributed Systems, Computer,Vol. 25, No. 12, Pp. 33-44, Dec. 1992.

5. N.Chandrakala Dr. P.Sivaprakasam,AnalysisOf Fault Tolerance Approaches In Dynamic Cloud Computing ©2013, IJARCSSE All Rights Reserved Volume 3, Issue 2,Pp 400-421,ISSN: 2277128X ,February 2013.

6. S. AoteAnd M. U. Kharat, A Game-Theoretic Model For Dynamic Load Balancing In Distributed Systems, In Proc. TheInternational Conference On Advances In Computing,Communication And Control (ICAC3 '09), Vol 4,Pp. 235-238,March 2009.