



INTEREST BASED CLUSTERING OF VIRTUAL MACHINES IN THE OVERLAY NETWORK OF CLOUD ENVIRONMENTS

D. Giridhar Kumar
Research Scholar
Dept. of CSE, SVUCE, India

D. Abdus Subhahan
Assistant Professor
Dept. of CSE, BVRIT, India

Abstract: In distributed environments, pay per usage of the user is allowed by Cloud Computing Networks. The usage may be either for computational or storage purpose which will be facilitated in the form of virtual machines in the virtual overlay. Provisioning of widely distributed resources efficiently to the user is the main objective. It in turn creates many challenges and issues like resource allocation problem, formulating Virtual Machines (VM), Load Balancing among VM's. In this paper we will see how Interest Based clustering of Virtual machines that reside over the physical overlay will enhance the performance. The request query from user may directly route to the cluster which can hit the query and internal Load Balancing among VM's traditionally.

Keywords: Cloud Computing, Resource Allocation, Load Balancing, Virtual Overlay

I. INTRODUCTION

Cloud Computing [1] is an emerging technology and can be seen as an innovation in different ways. Technically saying it's an advancement of computing, whose existence wave backs to the construction of the calculating machine. Cloud computing is being used widely because of its features like on-demand network access to a shared pool of computing resources (e.g., networks, servers, storage, applications, and services), less interaction, convenient to use.

Cloud computing can be defined as the delivery of hosted services over the Internet, such that accesses to shared hardware, software, databases, information, and all resources are provided to consumers on-demand by centralized management of resources and services. A cloud is one form of parallel and distributed system where there will be a collection of interconnected physical systems. All those are virtualized as one or more unified computing resources based on MoU (memorandum of understanding) between the service provider and the clients.

A. Cloud Components[3]

1. Client: Clients are nothing but the end users who are going to take the use of cloud services.
2. Provider: Cloud Service Provider is the owner of the particular cloud, who manages the requests from the user and provides services accordingly as per their requirements by taking nominal charges.
3. Datacenter: Cluster of actual resources (CPU, memory, storage, and so on) will formulate a Datacenter.
4. Virtual Machine: Virtual machine is a software implementation that executes programs as if they were actual physical machines.
5. *Physical Machine*: Actual machines (supplier of hardware like RAM, Processor) which exist in the underlay network are treated as Physical machines.

In cloud computing environments, resource allocation or load balancing takes place at two levels. First level is at, when there is a upload of an application to the cloud, the allocator assigns the requested client to physical machines, and the load balancer attempts to balance the computational load of multiple requests across the underlying physical network. Second level of load balancing is at, when an application raised several requests then those should be served by checking the number of instances from the same application.

II. RESOURCE MAPPING

In cloud computing, the objective of Resource Allocation/Mapping [2] is to utilize sufficiently the existing resources among the queried cloud clients over the internet. Resource allocation will get starved (infinitely waiting for resources) if no proper allocation strategy. Solution to such problem is within the hands of the service provider by proper usage of resource allocation schemes.

Resource Allocation Scheme (RAS) is all about proper allocation of existing physical resources, virtually among all the users who request for the service with the provider. It should be such that, the existing resources should serve the needs of the users. It will be done by considering what is the requirement and what type of resources are needed to service the user satisfactorily. An optimal Scheme should also consider the time when request raised and when it has been served such the delay in time should be minimum. An optimal RAS should mitigate the following criteria [10]:

- a) Resource contention is a situation that occurs when two or more clients requesting for same resources at a particular moment.
- b) If we have limited number of resources then Scarcity may occur.

c) Resource fragmentation: Even though we have adequate resources, it may not be allocated to clients due to isolation of resources.

d) Over-provisioning of resources is the situation where we are wasting resources by providing surplus to the client then its actual need.

e) Under-provisioning of resources is a situation where client is provided with resources less than the requested.

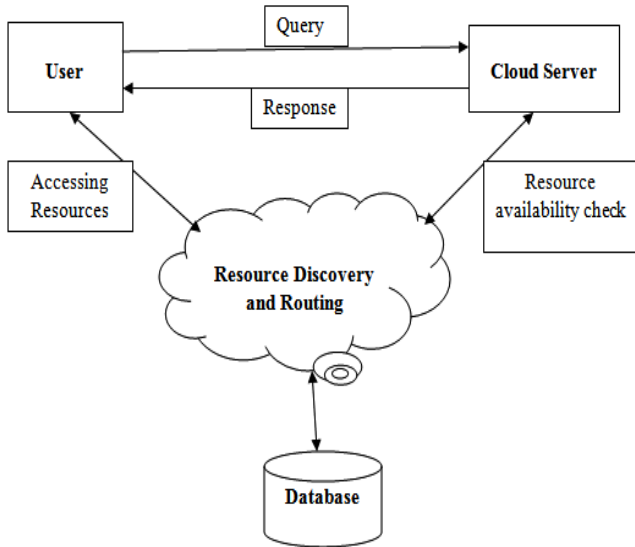


Figure 1. Resource Allocation Scheme

Cloud resources are categorized into two types; they are as physical resources and virtual resources. By using Virtualization concept, we use physical resources at distinct locations where the requests arise. Identification of the virtual resource for allocation is through the request parameters like hardware requirement (disk storage) or software requirement (processor, application software’s etc).

The provisioning request gets satisfied by mapping virtualized resources at VM’s to the actual resources at physical machines. The services may rely for either software or for hardware which will be served as per user’s requirement. Virtual machines are being used for attaining the scalability in cloud environments.

III. LOAD BALANCING

Load Balancing[6][9] is a method to distribute workload on the multiple computers or a computer cluster through network links to achieve optimal resource utilization for maximizing throughput and minimizing overall response time. Load Balancing is used for avoiding too much overload on the resources and dividing the traffic between servers and data. Data can be sent and received without maximum delay.

The main objective of load balancing is to mitigate the user’s waiting time for the service. We use Load Balancing Schemes [4] so as to balancing the work load on individual Virtual machine based on their capability.

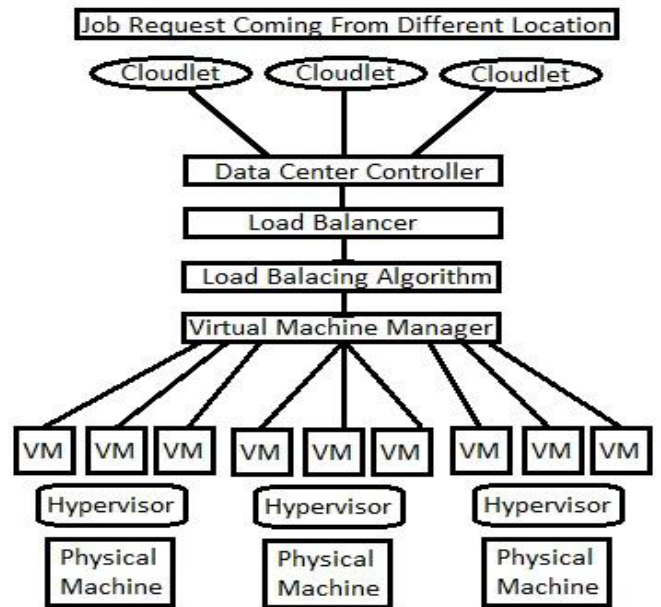


Figure 2. Load Balancing Mechanism

A. Goals of Load Balancing:

In particular Load Balancing Algorithms should strive to achieve some or all of the following conflicting goals

1. Minimizing Load Imbalance
Overhead associated due to movement of tasks, inter process communication should be reduced.
2. Responsiveness
There will be virtual machine migration, so algorithm should react to dynamically changing load distribution to yield good response to the user.
3. Efficiency and Effectiveness
Efficiency will be high if the performance gets enhanced similarly effectiveness is the measure of number of users that are being provided with resources in a fixed interval of time.
4. Decentralisation
As of Centralized Systems, there should not be single point of failure.
5. Resource Utilization
Utilization of resources should be maximum by making use of proper Resource allocation strategies.

IV. VIRTUAL OVERLAYS

A virtual overlay network [7] is a form of network virtualization which provides path between application software to underlying physical machines, with the help of hypervisors. Virtual overlay network is a partitioning between the virtual network and the underlying physical network hardware, enabling provisioning of virtual resources (storage and services) in the data center. That means based on the demand the VM’s will increase to support coming requests.

These overlay networks create several virtual networks that are logically independent but physically share the same underlying network. As the Virtual Machine migrate across the physical network, its included policies, security and VLAN's will be allowed by the controller to migrate as well.

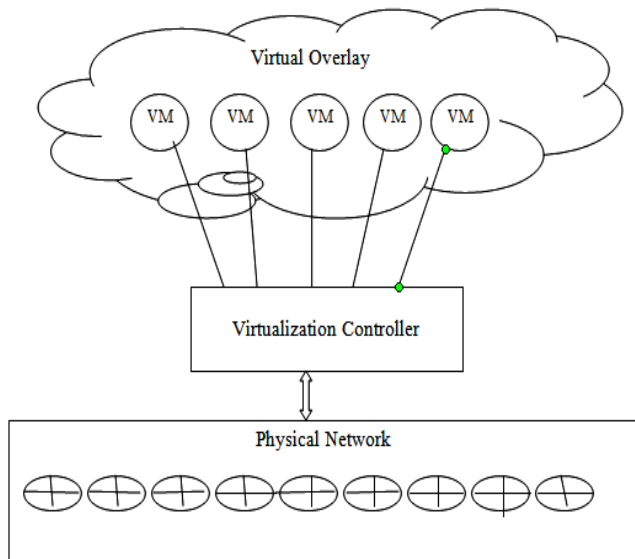


Figure 3. Virtual Overlays

A. Advantages of having Virtual Overlay Networks

In the case of large data center environments like Cloud Networks, virtual overlay network technology provides the following advantages:

1. Isolation of Traffic:
When there are multiple tenants/clients, cloud service providers have to isolate the traffic. It may be isolation of business unit traffic from production traffic in case of R&D Networks.
2. Provisioning of VM's:
There is a freedom for virtual machines to migrate from one location to another without considering about the underlying physical network. The policies, securities will also move along with VM's.
3. Scalability:
Due to usage of Virtual machines the network may expand to several thousand which is beyond the VLAN Capacity.
4. Network Independence:
As Virtual Overlay enables better migration n scalability at lesser costs dependency on the physical network reduced.

V. INTEREST BASED CLUSTERING

Virtual machines in cloud computing are those which provide the resources requested by the user. Clustering means grouping together, the similar entities. Clustering of VM's refer to the grouping of virtual machines which can serve similar kind of query requests. Usually the requests from the user may be for RAM, Bandwidth, Memory Storage, CPU

Cycles, Renting of Software's etc. So based on these parameters the clustering of virtual machines has to be done. Within each cluster there will be traditional Load Balancing mechanisms to monitor the load on each machine. Finally we will enable Interest Based communication channels among the clusters. It will be decided based on the number of requests being arised from the user for that particular combination of resources.

Example: If there are more number of requests for the combination of RAM and Memory Storage then a communication link is directly provided between Cluster of RAM sharing VM's to the Cluster of Memory Storage VM's.

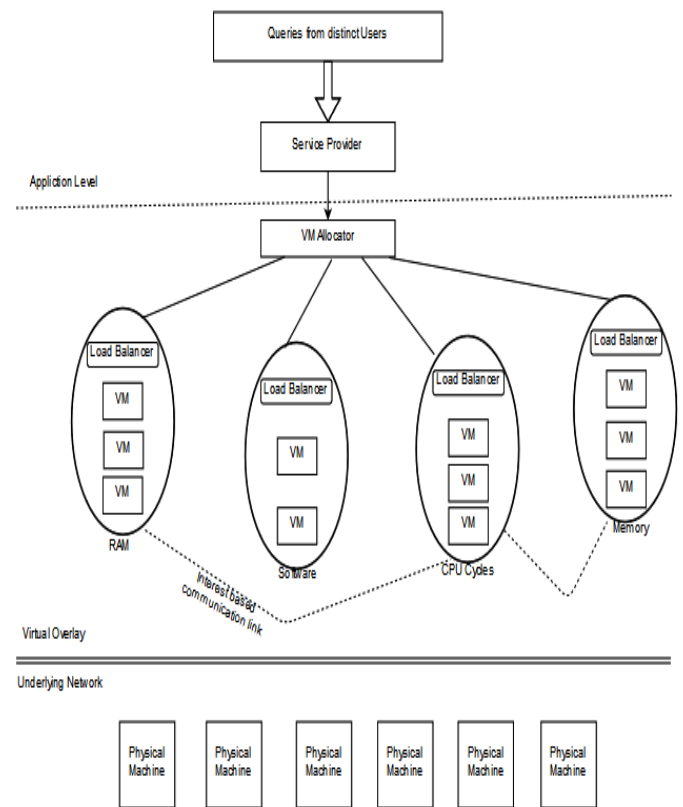


Figure 4. Interest Based Clustering of VM's

From the above figure it is clear that user may request the service provider for the resources through the application software at the application level. The service provider will implement some allocation strategies to find and allocate a cluster of virtual machines based on the type of the request by the user. We have used a load balancing scheme at every cluster level to balance load of that cluster. If the request is for combination of two or more different types of resources then by using interest based communication links they will be served quickly. All this process appears virtually as if done on a single machine.

VI. CONCLUSION

A virtual machine is a virtual form of computer hardware or software resources that are provided to the cloud users using a application software. Virtual machines will make the users feel as working on actual machines at the underlay network. We have given a short review of need for proper

resource allocation strategy and why load balancing is critical in cloud networks. The proposed Interest based clustering will enhance the performance by reducing the routing delay (by directly allocating the VM cluster based on the type of request and with internal load balancers). But here we are using distinct load balancers for distinct clusters which can be enhanced by a single level of load balancing schemes.

VII. REFERENCES

- [1] C. Papagianni, A. Leivadeas, S. Papavassiliou, V. Maglaris, C. Cervello-Pastor, A. Monje, "On the Optimal Allocation of Virtual Resources in Cloud-Computing Networks," *IEEE Trans. Computers*, vol. 62, no. 6, pp. 1060-1071, Jun. 2013, doi: 10.1109/TC.2013.31.
- [2] V. Vinothina, Dr. R. Shridaran, and Dr. Padmavathi Ganpathi, "A survey on resource allocation strategies in cloud computing," *International Journal of Advanced Computer Science and Applications*, 3(6):97--104, 2012.
- [3] Prof. Rupali M.Pandharpatte, "A Review: Resource Allocation Problem in Cloud Environment," *International Journal of Engineering and Technology (IJET)*, vol. 9 no. 3, pp. 1695-1700, jun-jul 2017, doi: 10.21817/ijet/2017/v9i3/170903038
- [4] Nusrat Pasha, Dr. Amit Agarwal, Dr. Ravi Rastogi, "Round Robin Approach for VM Load Balancing Algorithm in Cloud Computing Environment," *International Journal of Advanced Research in Computer Science and Software Engineering(IJARCSSE)*, vol. 4, no. 9, pp. 34-39, may 2014
- [5] J. Fan and M.H. Ammar, "Dynamic Topology Configuration in Service Overlay Networks: A Study of Reconfiguration Policies," *Proc. IEEE INFOCOM '06*, pp. 1-12, Apr. 2006, doi:0.1109/INFOCOM.2006.139.
- [6] Dharmesh Kashyap, Jaydeep Viradiya, "A Survey Of Various Load Balancing Algorithms In Cloud Computing", *INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH VOLUME 3, ISSUE 11*, pp. 115-119 NOVEMBER 2014
- [7] Navdeep kaur, Pooja Nagpal, "A Study on VMM and Resource Allocation Strategies in Cloud Computing Environment," *International Journal Of Engineering And Computer Science (IJecs)* ISSN:2319-7242 Volume 6 Issue 7 July 2017, Page No. 22116-22120 DOI: 10.18535/ijecs/v6i7.35
- [8] Mayanka Katyal, Atul Mishra, "A Comparative Study of Load Balancing Algorithms in Cloud Computing Environment" *International Journal of Distributed and Cloud Computing*, Volume 1 Issue 2 December 2013, pp. 5-14
- [9] Samarsinh Prakash Jadhav, Priya R. Deshpande, "Load Balancing in Cloud Computing", *International Journal of Science and Research (IJSR)*, Volume 3 Issue 6, June 2014, pp. 2282-2285
- [10] K.Delhi Babu, D.Giridhar Kumar, "Allocation Strategies of Virtual Resources in Cloud-Computing Networks", *Int. Journal of Engineering Research and Application* ISSN : 2248-9622, Vol. 4, Issue 11(Version - 5), November 2014, pp.51-55