Survey on Dynamic Resource Allocation of Virtual Machines

T. R. Abinaya[#1]
PG Scholar (CSE)[#1],
Sri Ramakrishna Engineering College,
Coimbatore,

R. S. Prabakaran[#2]
PG Scholar (CSE)[#2],
Sri Ramakrishna Engineering College,
Coimbatore,

T. Aarthi[#3]
PG Scholar (CSE)[#3],
Sri Ramakrishna Engineering College,
Coimbatore,

Abstract

Cloud computing is booming technology where they provide various services to the end users. Clouds can make to access applications and associated data from anywhere. Companies can rent their required resource from cloud like storage and other computational purposes so that their infrastructure cost can be reduced significantly. So we can make use of company-wide access to applications, mainly based on pay-as-you-go model. Because of this there is no need for getting licenses for individual products. One of the major disadvantage in cloud computing is related to optimizing the resource allocation. Because resource allocation is performed with the objective of minimizing the costs associated with it. There are many challenges of resource allocation are meeting customer demands and application requirements. So we focus on, various resource allocation strategies and their challenges are discussed.

***Keywords**- Cloud Computing; Cloud Services; Resource Allocation;*

Introduction

Cloud systems became a crucial computing platform of selection for medium and small-size enterprises and conjointly for the people, attributable to their cost-effective services. They integrate economies of scale in order that build them a lovely various to having Associate in Nursing in-house computing infrastructure. Main problems to be thought-about square measure with efficiency allocating the Virtual Machine for the specified resource. best virtual machine placement in clouds has received significant additional attention recently to facilitate cloud resources Associate in services provisioning on an on demand basis to users and multiple tenants. We tend to conjointly focus to dynamic variations of workloads, jobs and application flows to been focus. However the particular demand is often not known prior. Thus this makes reservation supported demand cloud services provisioning troublesome since it's impracticable to order the correct quantity of resources a priori.

Virtualization is the logical separation of the requested services and providing services to those requested service. So in particle they provide to run the various applications, operating systems and system services in various environments independently and with the specific physical machine. So we have to dynamically distribute the resources to the Virtual machines to complete the task with specific time efficiently.

Reasons for Resource allocation

- Resource allocation method helps to avoid overload in the system and also to minimize the number of servers used.
- Skewness- to measure the uneven utilization of a server.

1.Dynamic Resource Allocation Using Virtual Machines for Cloud Computing Environment [1]

The main aim of the algorithm in this paper:

- Overload avoidance – The availability of a PM should be satisfied the resource required by the Vm to run. Else the PM is overloaded this leads to decrease the performance of VMs.

- Green computing. PMs should minimized where they have to satisfy the requirements of VMs and also ideal PMs should turn off to avoid energy.

When there is natural transaction between two tasks we need to face the problems like resource requested by VMs. To avoid overload of tasks we should monitor the PMs to utilize the resource minimum where to use the resource later when the VMs requested

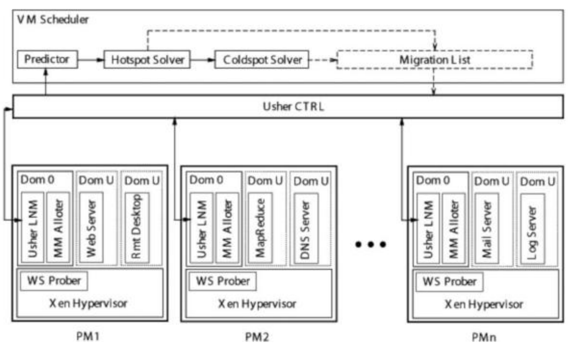


Fig:1 System Architecture

THE SKEWNESS ALGORITHM

So they introduce new concept of skewness to measure the unevenness in the utilization of resources on a server. Let n be the number of resources and r_i be the utilization of the ith resource.

So skewness of a server p as

$$skewness(p) = \sqrt{\sum_{i=1}^{n}\left(\frac{r_i}{\bar{r}} - 1\right)^2},$$

Where r is the average utilization of all resources for server p.

In general all resource performed well but thry consider only bottleneck resource By minimizing the skewness, we can combine different types of workloads and also to improve the resource utilization of servers.

2.Supplementary File: Dynamic Resource Allocation using Virtual Machines for Cloud Computing Environment

The skewness algorithm divide into three parts: load prediction, hot spot mitigation, and green computing. Let n and m be the number of PMs and VMs in the system. The number of resources like CPU, memory, I/O, etc. that need to be considered is usually a small constant

Hot spot mitigation

n_h - the number of hot spots in the system during a decision. Where they are ordered based on their temperature takes $O(n_h * \log(n_h))$. Next we have to order the VMs to run on it. To short the VMs we need time. To Place VMs with suitable destination it takes O(n). The overall complexity of this phase $O(n_h *n)$.

Green Computing

n_c -be the number of cold spots in the system during a decision run.

To sorting them based on their memory sizes it takes $O(n_c*\log(n_c))$. For each VM in a cold spot, it takes O(n) time to find a destination server for it.The overall complexity is $O(n_c *n)$. So they can get as high as O(n2)

Number of migrations

In the Hot Spot migration for each hot spot required at most one movement. In green computing, the number of migrations depends on number of VMs on the cold spots. So the total number of migrations is bounded by O(n).

3. Cost Efficient Datacenter Selection for Cloud Services[3]

This method helps to reduce the electricity cost for the data center at various location. They also consider network bandwidth. We propose to utilize statistical multiplexing to strategically bundle demands at different locations. The anti-correlation between demands effectively smooths out the aggregated bandwidth usage, thereby saving the bandwidth cost calculated by burstable billing methods that charge the peak bandwidth usage. We present an optimization framework that models the realistic environment and practical constraints a cloud faces.

Our main contribution in this paper is a general optimization framework for cost efficient datacenter selection that takes into account both electricity and bandwidth costs. Our framework is general in the sense that it models practical environments that a cloud operates in. The utility abstraction encompasses many performance considerations, including throughput, latency, as well as possible fairness criteria. The electricity and bandwidth cost constraints capture the two most important ongoing costs associated with the operation of datacenters, i.e. operation expense (OPEX) . Both the utility and electricity price are location dependent in order to realistically model the geographic diversity.

$$L_d = w_d^\intercal D.$$

$$L(w,\lambda,\delta,\nu) = \sum_d w_d^\intercal u_d + \lambda\left(A - \sum_d\left(w_d^\intercal \mu + \theta\sqrt{w_d^\intercal \Sigma w_d}\right)\right) + \delta\left(B - \sum_d p_d \cdot \alpha w_d^\intercal \mu\right) + \sum_i \nu_i\left(1 - \sum_d w_{id}\right),$$

where λ, &, and ⌫ are the Lagrange multipliers associated with the bandwidth usage, electricity cost, and work conservation constraints, respectively

4. Efficient allocation of virtual machine in cloud computing environment[4]

For the other two layers (i.e PaaS and SaaS), for their execution therefore we have focused on the IaaS. IaaS deliver computer infrastructure - typically a platform virtualization environment - as a service. Efficient scheduling of VMs instance request which meet user‟s requirements and improve the resource utilization increases the overall performance of the cloud computing environment. VM instance scheduling in IaaS is the one of the crucial cloud computing questions to address.

Suppose, „M‟ physical machines are available and their resource capacities given along memory, CPU and hard disk dimensions. There are „N‟ virtual machines to be placed. The requirements of these virtual machines are given along the dimensions of memory, CPU and hard disk. We have to find allocations of VMs on available physical machine that satisfies the VMs‟ resource requirements and increases overall the resource utilization. The objective of this research work is to introduce an algorithm for efficient VM scheduling in cloud computing in terms of resource utilization rate. The proposed algorithm is compared with other existing algorithm for VM instance allocation.

Flowcharts

FCFS VM allocation policy works as follows:

Cloud user are requests to allocate VMs. „h Cloud Broker which acts on behalf of the Cloud user submits the VMs request list. „h An event has been generated for the submission of the VMs request list to the data center. „h One by one VM in the series in the list has been passing to create VM on the host If not found host with enough of PEs required by VM then VM cannot be created. „h If found first host with less number of PEs in use and fulfils VM PE requirement: „X If host has enough RAM to fulfil the VM.s RAM requirement then place VM on host. „X Else, VM is not created on the host. This VM is again sent to create a new list of VMs. VM allocation using proposed algorithm works as follows:

1.Cloud user requests for VMs.

2.Cloud Broker which acts on behalf of the Cloud user

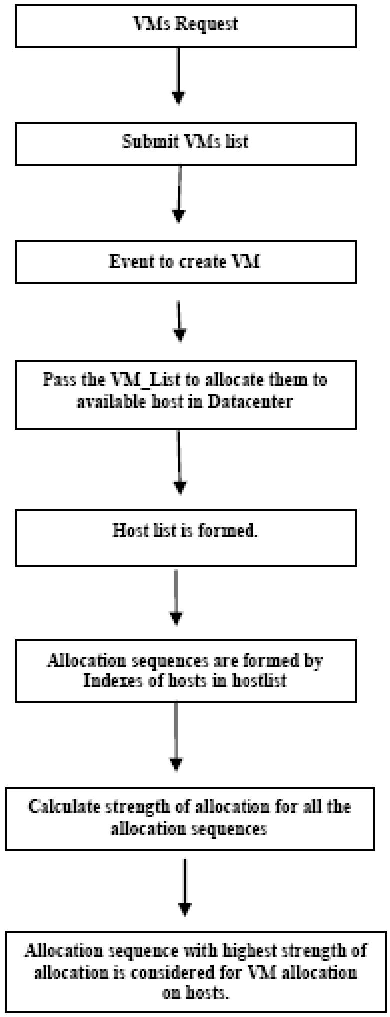
3.submits the VMs request list.

4.An event has been generated for the submission of the VMs request list to the data center.

The VM list is passed for the placement on the available host.

- Permutations of placement order of VMs on available hosts are considered.
- Strength of allocation is calculated for all the allocation sequence taken into consideration the VM„s parameters.
- Allocation sequence with the highest strength of allocation value is considered for

the VM allocation.



5.Energy Efficient Resource Management in Virtualized Cloud Data Centers [5]

The software system architecture is tiered comprising a dispatcher, global and local managers. The local managers reside on each physical node as a part of a Virtual Machine Monitor (VMM). They are responsible for observing current utilization of the node's resources and its thermal state.

The local managers choose VMs that have to be migrated to another node in the following cases:

• The utilization of some resource is close to 100% that creates a risk of SLA violation.

• The utilization of resources is low, therefore, all the VMs should be reallocated to another node and the idle node should be turned off.

• A VM has intensive network communication with another VM allocated to a different physical host.

• The temperature exceeds some limit and VMs have to be migrated in order to reduce load on the cooling system and allow the node to cool down naturally.

the system operation consists of the following steps:

1) New requests for VM provisioning. Users submit requests for provisioning of VMs.

2) Dispatching requests for VM provisioning. The dispatcher distributes requests among global managers.

3) Intercommunication between global managers. The global managers exchange information about utilization of resources and VMs that have to be allocated.

4) Data about utilization of resources and VMs chosen to migrate. The local managers propagate information about resource utilization and VMs chosen to migrate to the global managers.

5) Migration commands. The global managers issue VM migration commands in order to optimize current allocation.

6) Commands for VM resizing and adjusting of power states. The local managers monitor their host nodes and issue commands for VM resizing and changes in power states of nodes.

7) VM resizing, scheduling and migration actions. According to the received commands, VMM performs actual resizing and migration of VMs as well as resource scheduling. The developed algorithms have to meet the following requirements:

• Decentralization and parallelism – to eliminate SPF and provide scalability.

• High performance – the system has to be able to quickly respond to changes in the workload.

• Guaranteed QoS – the algorithms have to provide reliable QoS by meeting SLA.

• Independence of the workload type – the algorithms have to be able to perform efficiently in mixed application environments.

The VM reallocation problem can be divided in two: selection of VMs to migrate and determining new placement of these VMs on physical hosts. The first part has to be considered separately for each optimization stage.

The second part is solved by application of a heuristic for semi online multidimensional bin-packing problem

6.Heuristic Based Resource Allocation Using Virtual Machine Migration: A Cloud Computing Perspective [6]

HEURISTIC BASED VM PLACEMENT

The heuristic based VM migration scenario is partitioned as follows:

1. Determining when a physical server is considered to be overloaded requiring live migration of one or more VMs from the physical server under consideration.
2. Determining when a physical server is considered as being under loaded hence it becomes a good candidate for hosting VMs that are being migrated from overloaded physical servers.
3. Selection of VMs that should be migrated from an overloaded physical server. VM selection policy (algorithm) has to be applied to carry out the selection process.
4. Finding a new placement of the VMs selected for migration from the overload and physical servers and finding the best physical

The VM placement problem can be considered as a bin packing problem with variable been sizes and items, where bins represent the physical servers, items represent the VMs to be allocated, and bin sizes represent the available CPU capacities of those nodes. Since bin packing problem is NP-hard, to incorporate it into our solution we apply a modification of the Best Fit Decreasing (BFD) algorithm that uses no more than we take as an input the sorted list of VMs to be migrated in descending order of their current CPU utilizations and allocate each to a selected host that provides the least number of remaining processing capacity caused by the allocation. The host list is also sorted in decreasing order of their remaining capacity to ensure that a VM is allocated to a host that has enough resources for it with the least number of attempts. This ensure high resource utilization as the resources on the target host will not be idle.

The algorithm is presented below. It has the time complexity of nm, where n is the number of physical target hosts and m is the number of VMs that have been selected for migration.

Modified Best Fit Decreasing Algorithm:

1. Input: SortedlistofTargetHosts, SortedVmList, Output: VMs allocations
2. Foreach VM in SortedVmList do
3. minCapacity←MAX
4. allocatedHost←NULL
5. Foreach host in sortedListofTargetHosts do
6. If target host has enough resources for VM
7. Capacity←estimateCapacity(host,vm)
8. If Capacity<minCapacity
9. Allocated host←host
10. minCapacity←Capacity
11. If allocatedHost≠NULL then
12. Allocation.add(vm,allocatedHost)
13. Return allocation

Reference

[1] Zhen Xiao, Senior Member, IEEE, Weijia Song, and Qi Chen **"Dynamic Resource Allocation Using Virtual Machines for Cloud Computing Environment",** IEEE transactions on parallel and distributed systems, vol. 24, no. 6, june 2013

[2] Zhen xiao, senior member, ieee, weijia song, and qi chen, "**supplementary file: dynamic resource Allocation using virtual machines for cloud Computing environment**" IEEE transaction on parallel and distributed systems(tpds), vol. N, no. N, month year

[3] Hong Xu, Baochun Li henryxu, bli@eecg.toronto.edu Department of Electrical and Computer Engineering University of Toronto **"Cost Efficient Datacenter Selection for Cloud Services"**

[4] Sonam Rathore, Faculty of Engineering and Technology, MITS, Lakshmangarh, India "Efficient Allocation Of Virtual Machine In Cloud Computing Environment", International Journal of Computer Science and Informatics ISSN (PRINT): 2231 –5292, Volume-2, Issue-3, 2012

[5] Anton Beloglazov and Rajkumar Buyya Cloud Computing and Distributed Systems (CLOUDS) LaboratoryDepartment of Computer Science and Software Engineering The University of Melbourne, Australia **"Energy Efficient Allocation of Virtual Machines in Cloud Data Centers"** 2010 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing

[6] Ts`epoMofolo1, R Suchithra2 1,2MS (IT) Department, Jain University Bangalore, India **"Heuristic Based Resource Allocation Using Virtual Machine Migration: A Cloud Computing Perspective",** International Refereed Journal of Engineering and Science (IRJES) ISSN (Online) 2319-183X, (Print) 2319-1821Volume 2, Issue 5(May 2013), PP.40-45