A Survey on Energy Efficient Virtual Machine Placement in Cloud

Verma Megha[1] Jignesh Vania[2]
[1]M.E (Pursuing) [2]Faculty
[1,2]Department of Computer Engineering
[1,2]L.J.I.E.T., Ahmedabad, India

Abstract— Cloud computing is one of the most emerging technology of the world. Due to rapid increase in the number of data centers to provide the services to the users the power consumed and operational costs of them is going on increasing day by day. Due to high power consumption and continuous working the amount of CO2 emission is going on increasing which is adding up to the greenhouse effect. So there is a need to create an energy efficient system where the power consumption is less. Virtualization can be used to reduce the power consumption by data center. VM placement deals with the optimal choice of physical machine where it should be placed. Moreover the main aim for making it energy efficient is to transfer the load of fewer servers and switch off the idle servers. So in this paper a summary of various VM placement algorithms are discussed and energy conservation parameters are also discussed.

Key words: Green computing, VM placement, live migration, energy management

I. INTRODUCTION

Cloud computing is one of the most emerging fields of the world. It follows "pay as you go" model. This means that you only need to pay for the services which you use. In today's scenario it is very costly to build up your own applications, software, infrastructure, etc. so cloud providers rent the users their services and cost effective way to use them.

The demand for services is increasing day by day due rapid growth and this has led to the formation of large data centers. Moreover these data centers consume very large amount of energy and power, so we need to find a solution for reducing the power consumption. Apart from this the very high power consumption has led to the increase in carbon-dioxide components in the atmosphere. As a result of this the greenhouse effect has increased so we need to find some measures for decreasing the power consumption and environment friendly cloud computing.

The study mentioned in [1] shows that the total power consumed by data center was around 38 Giga Watt (GW) in 2012. By comparing the results with 2011 it was found to be 63% more consumption of power. According the McKinsey report [2], the authors of [3] have found out"The total estimated energy bill for data center in 2010 is 11.5 billion and energy costs in a typical data center double every five years". Hence it becomes very important to reduce power consumption.

II. GREEN COMPUTING

Green computing has gained focus in the last few years because of current energy optimization. Cloud providers do not focus on energy consumption. Almost 50% of the infrastructure remains unused. There are various energy saving techniques for green cloud computing to reduce cost, power consumption, energy consumption, CO2 emission, etc. Here I have focused on virtual machine placement strategy for reducing the power consumption.

The main focus is to use virtualization and migration techniques to find out where the VM (virtual machine) should be placed i.e. on which particular PM (physical machine). The ideal servers should be shut down for decreasing power consumption and their workload be shifted to other server.

The result will be energy efficient virtual machine placement in cloud algorithm.

The green cloud architecture [4] is as shown:

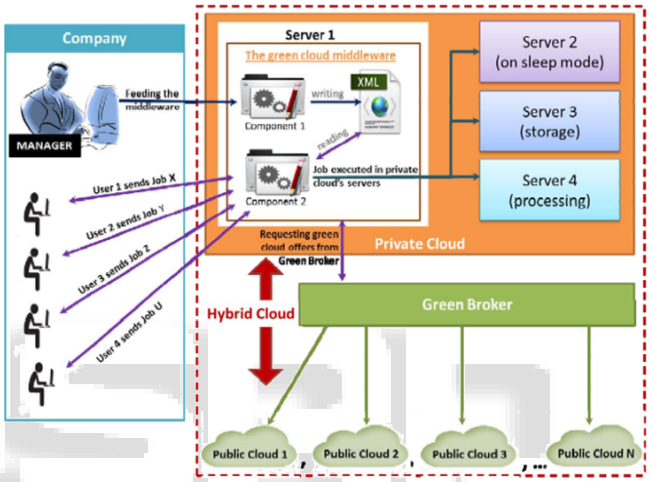

Fig. 1: Green Cloud Architecture [4]

As shown in figure, the client can be a company requesting for services. The green cloud middleware manages the whole scenario of allocating services to the cloud in such a way that the power consumption is less and more energy efficient framework. The user application and services build up the middleware. The private cloud is made up of few servers for storage, processing, etc. and the middleware. This architecture is a hybrid cloud which consists of both private and public cloud. The green cloud middleware requests the green broker to provide the green cloud.

III. VM PLACEMENT ALGORITHM

The researches have proposed [5], VM placement is the process of deciding where the virtual machine should be placed. It chooses the proper physical machine to host the virtual machine. VM decision has to be made at two places: First, when the VM is created, the care has to be taken that the VM is placed on the proper physical machine which can satisfy the needs (CPU utilization, memory, network, disk space) of the virtual machine. Second is when the VM is migrated, the care has to be taken that the target machine can take up the workload easily and the time and cost of migration is not more.

There are various strategies for virtual machine placement [6] which are described as below:

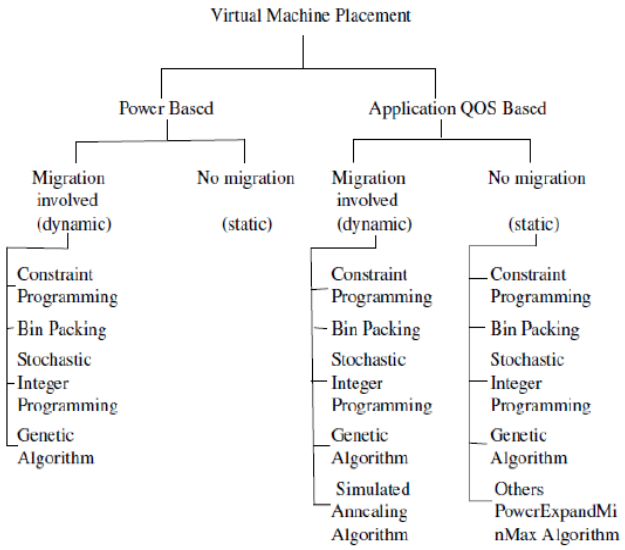

Fig. 2: Classification of VM Algorithm [6]

The VM placement algorithm can be classified into two basic approaches.

Power based approach is gaining its importance day by day due to the increase in cost of power and increasing demands for power consumption for computation and cooling of resources. The mapping of VM to PM is done in such a way that the servers are capable of being utilized to full extent and the idle servers can be switched off. This helps in improving the efficiency.

Application QOS Based Approach focuses on maximizing the quality of service. There is a constant monitoring of virtual machine and implementing policies for workload. This approach can lead to better utilization and savings.

The placement problem is non-deterministic hence we discuss few approaches.

- Constraint Programming,
- Bin Packing Algorithm,
- Stochastic Integer Programming,
- Genetic Algorithm.

A. *Constraint Programming:*

It is used for the combinatorial search problems where the solution should satisfy certain constraints and automatically manage the virtual machines. The basic steps include first of all declaration of domains of variables. Then we need to decide the constraints on declared variables and search for the domain. VM placement here is considered as two-step process. It depends on local decisions for each application environment. Then the local decisions of all applications are given as input to the global decisions. The main aim is to maximize the global utility function. VM problem as constraint programming can be explained as: *Constraint:* Number of available physical machines and maximizes the global utility function. *Model:* Find VM allocation vectors. This type of approach is useful when we have input data and we know the demands of VM. As the number of constraints increases more time is taken to reach the optimal solution.

B. *Bin Packing Algorithm:*

Here the various physical machines are assumed to be as bins and the virtual machines as different objects which have to be placed on the bins. The best efforts as made so that minimum numbers of physical machines are on at a same time. It has been summarized into three steps: firstly we need to study the past demands and form a pattern based on it, then forecast future demands based on the pattern observed earlier and at last map the virtual machines on the particular physical machines. This is called MFR (measure-forecast-remap).

This approach is mostly used for dynamic VM placement in the situations where demands are highly changing. As it is heuristic based approach it is difficult to get optimal solution but you can get a good solution in predefined amount of time. The dynamic placement reduces the number of PMs to half. It is helpful in the situations where PMs have same memory and configurations.

C. *Stochastic Integer Programming:*

There are few conditions where the actual demands are not known but there are few ways in which the demands can be estimated. Based on this the mapping of virtual machine to physical machine is done in three steps: Reservation- here the demands of users are not known and resources are allocated. Utilization- all the resources which have been reserved are used. On demand- there are certain situations when the reserved resources become scarce and demands increase, in such a scenario resources can be made available on demand basis.

This works best when the prices of resources and future demands are not known but somehow we can know or calculate the probability distributions. It can be helpful in the situations where there are two or more unknown cost dependent factors.

D. *Genetic Algorithm:*

It is a heuristic based search technique useful in the situations where objective functions change dynamically. It starts with initialization of solution and keeps on applying genetic operators on it till we get optimal solution. A genetic algorithm typically requires a genetic representation of the solution domain and fitness function to evaluate the solution domain.

This is generally used when there are static placements and demands don't vary much.

Hence depending on the applications and type of demands one can use any of the above algorithms to reduce the cost and power consumed.

IV. VM MIGRATION

VM migration is defined as a process where the virtual machine is moved transparently from one physical machine to another. In such a process the application or virtual machine on the host is active during the transfer. Here the memory, data storage, network elements are transferred from one machine to another. Basically there are two ways of transferring memory from one physical machine to another which are discussed as below.

A. *Pre-Copy Memory Migration:*

It works in two phases:

1) Warm-Up Phase:

Here in all the pages are copied on from source machine to the destination machine. During this transfer process the

virtual machine is still active. If some changes are made to the pages which have already been copied then such pages are resent to the destination machine. These pages will be recopied till the page recopied rate is not less than the page dirty rate.

2) Stop-and-Copy Phase:

Once the warm-up phase is completed the execution of source machine is stopped and the remaining dirty pages are recopied. After that the working is resumed at the destination machine. The time between the stopping of work on source machine and resuming it on the target machine is known as downtime and special efforts are made to reduce it so time is not wasted between executions.

B. Post-Copy Memory Migration:

In the post-copy migration the first step is to stop the execution on the source machine. The time when the execution is suspended some part of the source content (CPU registers, memory or pages) is transferred to the destination machine and the working is then resumed and started at the target machine. After the execution is started again other remaining content (memory, etc.) is transferred from the source to target machine. If someone tries to access the content/page which has not yet been transferred, at that time the page fault occurs and the content is unavailable. This sort of continuous situation can degrade the performance and affect the working.

V. EVALUATION METRICS FOR VM PLACEMENT

It is very important to study various parameters that affect the efficiency of the system. The following parameters add up to calculate the efficiency of VM placement algorithms and have been summarized in [5].

A. Energy Consumption:

As observed in [7], the total energy consumed is very important concern depending on which the total power consumption is calculated. Energy consumption is calculated for almost all VM placement and it affects the decision so proper care has to be taken to calculate it. As the power consumption increases the CPU utilization also increases. The power consumption is calculated from the idle state to fully operational state. The relationship can be shown as in (1)

$$P(u) = P_{idle} + (P_{busy} - P_{idle}) * u \qquad (1)$$

Where P(u) is the estimated power consumption
u is the current CPU utilization
P_{idle} is the power consumed by ideal server
P_{busy} is the power consumed by the server when it is fully utilized

B. Total Migration Time:

As observed by the authors in ([8], [9], [10]) the total migration time depends on many factors. The first one can be considered to be the amount of memory which has to be transferred from the source to the destination machine and the available network bandwidth. The relationship can be shown as

$$T_{mig} = V_{mem}/b \qquad (2)$$

where T_{mig} is the total migration time
V_{mem} is the amount of memory to be transferred
b is the available bandwidth

In the live migration the transfer of content is done in many phases. First of all the whole RAM is sent and after that the dirty pages are sent and recopied until there is consistency between the source and destination machines. The page dirty rate decides the number of pages that has to be resent whereas the VM memory usage pattern decides how many pages are to be resent. Consider the page size and previous round duration. The relationship for a single round i is as shown

$$V_{mig,i} = \begin{cases} V_{mem} & \text{if } i=0 \\ d*1*t_{i-1} & \end{cases} \qquad (3)$$

Where $V_{mig,i}$ is the amount of memory transferred in one round
d is the page dirty rate
is the page size
t_{i-1} is the duration of previous round

The total migration time is determined by the sum of the memory migrated during each round i , if the algorithm works for n rounds and finally ends with a stop and copy phase then the total migration time for the entire memory is shown as

$$V_{mig} = \sum_{i=1}^{n+1} V_{mig,i} + V_{mem}/b \qquad (4)$$

C. Downtime:

It has been seen in [11], that during the migration the VM is suspended during stop and copy phase for a time interval and the remaining dirty pages are transferred. There is a time interval may be of few milliseconds or seconds between the process of stopping the execution and resuming it at the other end. This time interval is known as the downtime and it should be as small as possible so that we don't incur any delay in synchronization. Hence the downtime can be shown as

$$T_{down} = (d*1*t_n)/b \qquad (5)$$

Where d is the page dirty rate
is the size of page
b is the link speed
t_n is the duration of n^{th} pre copy

VI. COST OF VM MIGRATION [3]

The authors in [12] made an effort to categorize the migration costs as below:

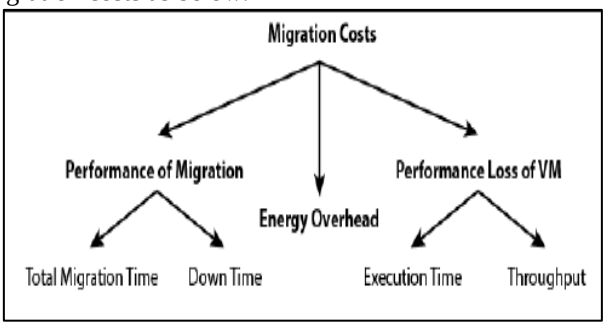

Fig. 3: Classification of Migration Cost [12]

The first important parameter affecting the cost of migration is the performance parameters. The *total* migration time is the total time taken when the migration process starts at the source machine until it is completely resumed and starts operating at the target machine. The other parameter is the downtime which is the time taken in the stop and copy phase wherein the execution is stopped for a particular duration of time.

The second parameter is *energy overhead.* This depends on the CPU utilization and increases with the consumption of energy. We need to take care of the energy consumed.

The third and important parameter is the performance loss during VM migration. The task is to find the dirty pages and send them. This process of finding the dirty pages requires a supervisor who takes care of all this and other resources are also utilized which increases the cost. The *execution time* is the total time taken for executing the tasks and *throughput* is the number of processes which can complete in the given time interval.

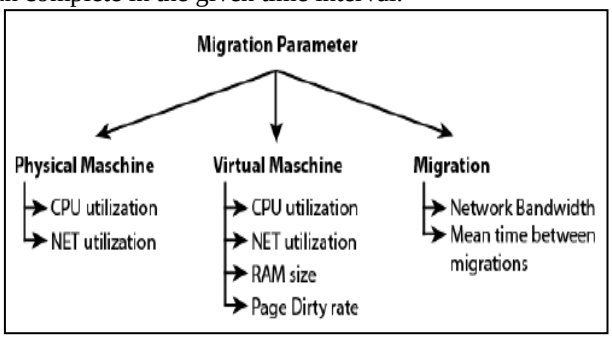

Fig. 4: Classification of Migration Parameters [12]

There are various parameters affecting migration. The physical machine related parameters include the resource utilization of source and target hosts. The current load of system affects the migration costs as this places an extra load on the CPU because of the main memory. The network elements like switches, network interface card, etc. add up to the cost. The virtual machine parameters depend on CPU utilization and network utilization. Other than this it depends on the RAM size which decides how many dirty pages are there i.e. page dirty rate. The migration related parameters refer to the average network bandwidth which is available during source to target transfer and second defines the meantime between migrations.

VII. CONCLUSION

This paper presents few approaches related to VM placement and the evaluation metrics and cost factors affecting VM migration. Due to increasing demands the energy saving techniques have gained popularity. The virtualization helps in this and idle servers can be shut down. Apart from this the source and destination machine can be identified for migration. Hence an effective use of resources can be helpful in achieving energy aware green computing. However the approach depends on the type of applications and demands.

REFERENCES

[1] "Taxes and incentives for renewable energy", KPMG international, June 2012.
[2] J. Kaplan, W. Forest, and N. Kindler, "Revolutionizing data center energy efficiency," McKinsey Company Tech Rep,p. 15, July 2008.
[3] Bharti Wadhwa, Amandeep Verma, "Energy Saving Approaches For Green Cloud Computing: A Review", in IEEE Conference on Engineering and Computational science(RAECS),2014, pp. 1-6
[4] Mohammad Naiim Hulkury, Mohammad Razvi Doomun, "Integrated Green Cloud Computing Architecture", IEEE International Conference on Advanced Computer Science Applications And Technologies (ACSAT), 2012, pp. 269-274
[5] R. Ranjana and Dr.J.Raja, "A Survey on Power Aware Virtual Machine Placement Strategies in a Cloud Data Center", in IEEE International Conference on Green Computing, Communication and Conservation of Energy (ICGCE), 2013, pp. 747-752
[6] Anjana Shankar, Virtual machine placement in computing clouds, Technical report, April 2010.
[7] X. Fan, W. D. Weber, and L. A. Barroso, "Power provisioning for a warehouse-sized Computer", in Proceedings of the 34th Annual International Symposium on Computer Architecture (ISCA 2007). ACM New York, NY, USA, 2007, pp. 13–23.
[8] S. Akoush, R. Sohan, A. Rice, A.W. Moore and A.Hopper, "Predicting the Performance of Virtual Machine Migration" in The 18th Annual Meeting of the IEEE International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems (MASCOTS 2010), Miami Beach, Florida, 17-19 Aug-2010, pp. 37-46
[9] Y. Kuno, K. Nii and S. Yamaguchi, "A Study on Performance of Processes in Migrating Virtual Machines", in 10th Intenational Symposium on Autonomous Decentralized Systems (ISADS), 23-27 March2011,pp.567-572.
[10] F. Salfner, P. Troger and A. Polze, "Downtime Analysis of Virtual Machine Live Migration", in: The Fourth International Conference on Dependability (DEPEND 2011), French Riviera, France, August 21-27,2011, pp. 100-105.
[11] H. Liu, C.-Z. Xu, H. Jin, J. Gong, and X. Liao, "Performance and energy modeling for live migration of virtual machines", in The 20th International Symposium on High Performance distributed Computing (HPDC'11). San Jose, California, USA, 8-11 June, pp. 171-182. 100-105.
[12] Anja Strunk, "Costs of Virtual Machine Live Migration: A Survey", in IEEE Eighth World Congress on Services, Honululu,2012,pp.323-329

