Middle-East Journal of Scientific Research
24 (Special Issue on Innovations in Information, Embedded and Communication Systems): 236-242, 2016
ISSN 1990-9233; © IDOSI Publications, 2016
DOI: 10.5829/idosi.mejsr.2016.24.IIECS.23166

Energy Aware Virtual Machine Consolidation in Conjunction with Qos for Green Cloud Computing

[1]M. Juliot Sophia and [2]Vejey Subash Gandyer

[1]Department of Computer Science and Engineering,
PG Scholar, KCG College of Technology, Chennai, India
[2]Department of Computer Science and Engineering,
Assistant Professor, KCG College of Technology, Chennai, India

Abstract: Today's digital world relies on cloud computing to achieve its business goals. Dependence on cloud computing solutions and their infrastructure results in increased energy consumption by the data centres, which in turn impose major impact on our ecosystem. Number of energy efficient solutions are introduced to minimize the effects of cloud computing. Green computing is one of the best energy efficient solution. It is an innovative way on how technology and ecology converge together. In this paper, we discuss about the design and implementation of a distributed system which runs on the basis of algorithms such as Modified best fit decreasing(MBFD), Earliest Dead line first (EDF) and the bio inspired hybrid ACO–GA algorithm. The system proposed here can be used in data centres to increase their energy efficiency.

Key words: Cloud Computing · VM Consolidation · Virtual Machine · Physical Machine · Energy Consumption · QOS

INTRODUCTION

Clouds are virtualized data centres. Cloud applications are offered as services on subscription basis. They require high energy usage for their operation [1]. This results not only in higher operational cost but also in tremendous amount of carbon dioxide (CO_2) emissions and global warming. Green cloud computing is used to counter balance carbon dioxide emission in the world with, limited energy resources and ever rising demand for more computational power. Data centres basically works on the basis of virtualizing the computer resources such as processors, memory and I/O devices. The applications of data centres have dynamic needs which makes the capacity management and demand forecasting as a complex task. This is the main challenge faced by data centres and it can be rectified with the help of energy-aware resource management, such as dynamic server provisioning and virtual machine consolidation.

A. Dynamic Server Provisioning Approaches [2]: Energy can be saved by using reduced amount of resources needed to satisfy the workload requirements. Therefore, unnecessary servers are switched off or put into lower power mode when the workload demand decreases. Similarly, when the demand increases additional servers are switched on or put back into high power mode.

Virtual Machine Consolidation: Sharing of Physical Machines (PMs) among various individual systems called Virtual machines (VMs), where each VM can run number of application tasks. The sharing of the PM resources among VMs can be effectively handled by Virtual Machine Monitor (VMM). It allows live VM migration and consolidation to pack VMs on limited number of PMs and so reduce the energy consumption by efficient placement of VMs on PMs. In Our paper we are proposing a distributed system architecture required in data centres to reduce the power utilization.

Energy Consumption by Data Centres: A Cloud data centre comprises of many hundreds or thousands of networked computers with their corresponding storage and networking subsystems, power distribution equipment, conditioning equipment and cooling infrastructures. Due to this, data centres can consume

Corresponding Author: M. Juliot Sophia, Department of Computer Science and Engineering,
PG Scholar, KCG College of Technology, Chennai, India.

massive energy and emit large amount of carbon dioxide. It has been clearly observed that energy consumption is not only because of servers and storage systems but also due to the cooling equipment of data centres. Lightning, power distribution, electrical equipments such as UPS are the other factors of energy consumption. Majority of power usage in a data centre is because of other purposes than actual IT services, so proper care should be taken while designing a data centre in such a way to ensure reduction in carbon footprint.

Methods to achieve Energy efficiency Target: Following methods are in use to attain energy efficiency

- Make use of technically advanced IT equipments.
- Proper sizing of UPS and other power distribution equipments which helps in minimization and elimination of unwanted AC/DC conversion.
- Following a hot/cold aisle configuration which involves the placement of servers in racks and rows to balance the heat generation.
- Consolidation and virtualization of servers are recommended to attain maximum utilization.
- Perfect cooling system and usage of outside air for cooling the data centre.

Consolidation and Virtualization: The most useful method of acquiring energy efficiency is obtained by consolidating and virtualizing the servers. In a data centre, all the servers are not fully utilized, mostly 90% of the servers are idle, performing nothing but consuming huge amount of heat and energy, so switching off the idle servers can save power and emission of heat and carbon di oxide. Placement of virtual machines should be efficiently done to reduce energy consumption. Following are the two effective methods to reduce energy consumption.

Consolidation: The approach in which multiple workloads run with single OS.

Virtualization: The approach used to reduce number of physical machines and can run any OS and the number of OS running will not be reduced. Virtualization is easy to implement than consolidation. It occurs by logically accommodating multiple virtual machines in to a physical server by sharing its hardware resources alone. Each VM can run its own operating system and applications. Hypervisors also called as virtual machine managers are used to allocate and control the VMs on Pms.

Techniques Used in Data Centers:

- Hibernating or turning off the idle servers
- Dynamic scaling the operating frequency/Voltage (DVFS) for underutilized servers
- Live VM Migration
- Running servers at power efficient rate
- Use power efficient servers to serve more number of requests.

This Paper Mainly Focus on Live VM Migration: Figure 1 depicts overall view of live VM migration to consolidate VMs. Some of under loaded PMs can be released for termination. Deciding when it is best to reallocate the VMs from an overloaded or under loaded PM is an important aspect of resource utilization and QOS.PM which doesn't have any VMs running on it can be switched off [3, 4]. Reducing number of PMs running can reduce the power utilization.

Service Level Agreement (SLA): Methods used to achieve energy efficiency should not affect Quality of service (QOS) requirement of the client. Following Service Level Agreement (SLA) is necessary to derive the QOS as expected. SLA is an interleaved performance of people, technology and process. Service providers should follow SLA very strictly in a way to satisfy the customers. Energy efficiency method selected should not degrade the data centre by violating SLA. Thus, an adequate care should be taken to select the methods for achieving energy efficiency.

Related Work:

Jian Cao *et al.*, proposes a paper "*Energy Efficient Allocation of Virtual Machines in Cloud Computing Environments Based on Demand Forecast*" which is ***u***sed to model time-varying demands or user behaviours of some seasonal pattern by the help of following methods and algorithms

Holt - Winters' Exponential Smoothing Method : It works on the basis of forecasting varying demands of data centres periodically and allocating VMs accordingly to reduce power consumption [5].

Knapsack Algorithm: Allocate VMs among PMs with higher resource provisioning under the conditions of limited energy consumption

Self-optimizing Module: Update values of parameters in forecasting mode and determine reasonable forecast period to set a threshold for forecast error.

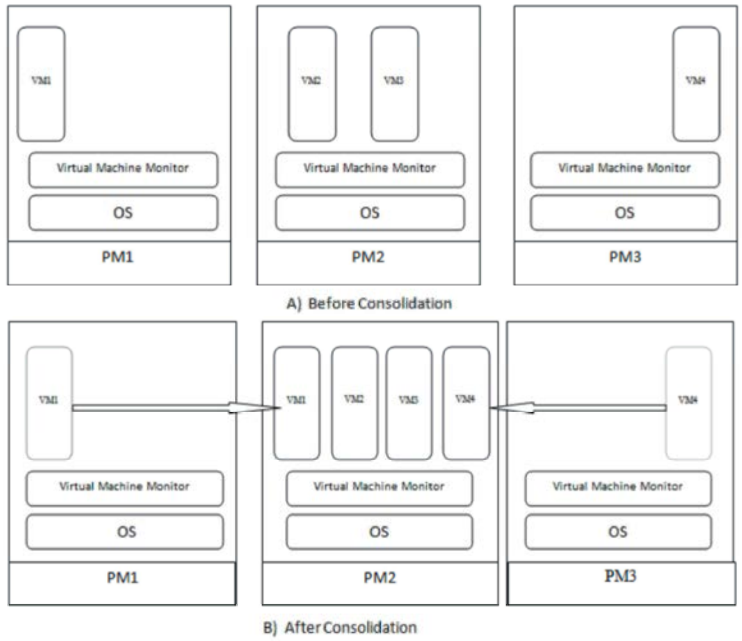

Fig. 1: View of Live VM Migration

Young Choon Lee *et al,* Introduces New Efforts in the Paper *Energy Efficient Utilization of Resources in Cloud Computing Systems* to reduce power consumption by idle machines, By the way of putting computer resources into sleep/power-saving mode, to increase resource utilization and reduce energy consumption [6]. Following two algorithms are implemented in this paper,

Energy Conscious Task Consolidation (ECTC): Calculates energy consumption of a task and subtracts the idle consumption if other tasks that are allocated on this physical machine at the same time.

Maximum Rate Utilization (Maxutil): Aims to utilize available resources as much as possible by choosing the solution where average utilization during the task execution is the highest.

ZhenXia *et al*., in the paper "*Dynamic Resource Allocation Using Virtual Machines for Cloud Computing Environment"* implements an automated resource management system to achieve good balance between Overload avoidance and Green computing [7]. It works on the basis of the Skewness algorithm which is used to quantify unevenness in utilization of multiple resources on a server.

Fahimeh Farahnakian *et al*., introduces a distributed system architecture in the paper "*Using Ant Colony System to Consolidate VMs for Green Cloud Computing"* to perform VM migration dynamically in order to improve resource utilizations of PMs and to reduce power consumption [3]. It uses following algorithms and cloud sim tool

Best fit Decreasing (BFD): Sort all virtual machines (VMs) in decreasing order based on the CPU requirements. After sorting, VMs at the top is selected and placed on an already used PM that has minimum CPU capacity. During unavailability of resource on used server(s), VMs are placed on a new PM that has minimum CPU capacity.

Ant Colony System (ACS): A set of agents mentioned as *ants* cooperate to find best solution by exploitation & exploration. Ants communicate using pheromone deposit on the edges (stigmergy) which results in local Pheromone trial & global Pheromone trial [8, 9].

Cloud Sim: A discrete event simulator for implementation and evaluation of resource provisioning and VM consolidation techniques for different applications.

Jashweeni Nandanwar *et al*. in the paper " *Aggregation of EDF and ACO for Enhancing Real Time System Performance"* proposes an adaptive algorithm which combines both static and dynamic algorithm. The goal of this adaptive scheduling algorithm is to show the switching between the scheduling algorithms and to reduce failure by increasing the performance [10]. This paper uses following algorithms,

Table 1: Comparison of algorithms

Algorithm	Advantages	Limitations
Knapsack Algorithm	Frequency of switching on/off hosts can be reduced to reduce total energy consumption	Inefficient when demands are unusually high or low
Energy Conscious Task Consolidation and Maximum Utility	These heuristics have effective energy-saving possibilities with better resource provisioning	Migrated tasks may hinder consolidation of new arriving tasks, resulting in more energy consumption
The Skewness algorithm	Achieves both overload avoidance and green computing	Need to predict the future resource needs of VMs and impact on application performance due to live VM migration
Best Fit Decreasing	Best suited for worst case and for average uniform case	Focus is only on space consumption and not on energy consumption
Ant Colony System	Aims to concentrate the search in regions of the search space containing high quality solutions	Takes more number of iterations to complete the task
Adaptive Earliest Deadline first and Ant Colony Optimization	Schedules more process by maximizing CPU utilization and minimizing the energy Consumption. Useful when future load is unpredictable	Increase in memory usage when the load increases butswitching between the algorithm makes the code complex
Hybrid Genetic and Ant Colony Optimization	ACO is a temporary solution generator that will be improved by implementing GA operators iteratively until the stop criteria is reached	Shows some sort of insignificancy in small amount of data and optimization of certain parameters is required to get best result

Earliest Deadline First: Handles under loaded condition. It arranges the task in a priority queue by placing the task which is closest to the deadline at the top of the queue.

Ant Colony Optimization: Handles the overloaded condition which is a branch of swarm Intelligence. It requires more processing time while compared with EDF [8, 9].

Adaptive Scheduling Algorithm: Switched between EDF and ACO based on current load.

Zainudin Zukhri *et al*., in the paper "*A hybrid optimization algorithm based on genetic algorithm and ant colony optimization*" uses THE SURVIVAL OF THE FITTEST idea to provide a different yet efficient searching technique which explores selected possible solution to obtain good result [11].

Genetic Algorithm (GA): This algorithm is based on natural evolution which explores selected possible solution to obtain good result. It is used to minimize the search space by selecting the best fit.

Ant Colony Optimization (ACO): Output from GA is used as the input of ACO algorithm.

Hybrid GA and ACO: It is the combination of Genetic Algorithm and Ant colony Optimization algorithm.Both GA and ACO algorithms are combined to get best solution.Hybridization is applied to some parameters and variables of GA or ACO.

Benchmarking: Based upon the comparison study of algorithms, we have decided to make use of following algorithms in our proposed system

Algorithms:

Modified Best Fit Decreasing Algorithm: All the VMs are sorted in decreasing order of current CPU utilizations and each VM is allocated to a host which provides the least increase in power consumption. This allows to use more power efficient PMs first.

Earliest Deadline First Algorithm:

- It is a priority driven algorithm where higher priority is assigned to the task with earliest deadline.
- It is based on dynamic priority assignment as the priority of the task is assigned as the task arrives. When a new task arrives it is inserted into a queue of ready tasks sorted by their deadlines.

Ant Colony Optimization Algorithm:

- This algorithm is inspired by the way the ants seek shortest path from the food source to their colony by the help of pheromone deposit.
- It is best known for search space exploration by acting both concurrently and independently.

Genetic Algorithm:

- This algorithm works on the basis of survival of fittest idea and follows the sequence of actions as follows : generating the initial population, evaluation, selection, crossover, mutation and regeneration
- It is mainly suitable for many optimization problems since it doesn't require any prior knowledge of the problem to be solved.

Proposed System: A cloud data centre comprises of number of physical machines with different configurations. Client request are satisfied by allocating the virtual machines on these physical machines based upon the requirement. This allocation should be done without degrading the quality of service. Our proposed system presents a distributed architecture which can be used in data centres to increase energy efficiency without missing the clients' requirements. The proposed system consists of global agent where our adaptive algorithms are implemented. Initially the VMS are allocated on PMS using MBFD algorithm. During under loaded condition the architecture choose the Earliest dead line First (EDF) algorithm and in case of overloaded condition implementation of hybrid ACO-GA algorithm takes place. Under loaded and overloaded condition can be identified by setting up a threshold. If more than 50 percent of overall PMs are used, it is overloaded else it is under loaded.

Implementation of EDF: VMs are allocated using EDF only if they pass the schedulability test. A VM is allocated using EDF, if it satisfies the condition that the total processor utilization (Ui) due to the allocation is less than 1.

Schedulability test can be defined as,

$$U = \sum_{i=1}^{n} \frac{C_i}{T_i} \leq 1$$

C_i - Worst case computation of n process
T_i - Inter arrival periods (equal to the deadline)

Implementation of hybrid ACO-GA: The objective is to find a migration plan which can be stated as,

$$f(M) = |P_S|^{\gamma} + \frac{1}{|M|}$$

M - Migration plan
P_S - Set of physical machines that will be made in sleep mode due to the implementation of Migration plan M.PMs only without any running VMs can be in sleep mode.
γ - Parameter to determine relative importance of $|P_S|$w.r.t |M|

Hybrid ACO-GA Can Be Explained in Following Steps: Step 1: Paths' Finding: The ants search for the best PM to migrate the Vms from either underutilized or over utilized PMs. An ant K chooses a tuple s to traverse next by applying the pseudo-random-proportional rule

$$s = \begin{cases} \arg\max_{u \in T_k} \left\{ [\tau_u] \cdot [\eta_u]^{\beta} \right\}, & \text{if } q < q_0 \\ S, & \text{otherwise,} \end{cases}$$

τ - Amount of pheromone
η - The heuristic value associated with particular tuple
β - Parameter to determine the relative importance of the heuristic value with respect to the pheromone value.
T_{RC}T - The set of tuples that remain to be traversed by ant k
q ∈ [0,1] - Uniformly distributed random variable
q_0 ∈ [0,1] - Parameter
S - Random variable

It is selected based on probability distribution given in following equation, where the probability of an ant k to choose tuple s to traverse next is defined as,

$$p_s = \begin{cases} \frac{[\tau_s] \cdot [\eta_s]^{\beta}}{\sum_{u \in T_k} [\tau_u][\eta_u]^{\beta}}, & \text{if } s \in T_k, \\ 0, & \text{otherwise.} \end{cases}$$

τ_s - Amount of Pheromone
η_s - Heuristic value associated with particular tuple

$$\eta_s = \begin{cases} (|C_{p_{de}} - (U_{p_{de}} + U_v|_1)^{-1}, & \text{if } U_{p_{de}} + U_v \leq C_{p_{de}} \\ 0, & \text{otherwise,} \end{cases}$$

C_{pde} - The total capacity vector of the destination PM *pde*,
U_{pde} Used capacity vector of *pde*,
U_v - Used capacity vector of the VM v in tuple s

Local Pheromone Trail Update: An ant reaches a tuple while applying a migration plan where the following local pheromone trail update can be used. It is defined as

$$\tau_s = (1-\rho).\tau_s + \rho.\tau_0,$$

ρ - Pheromone decay parameter in the global updating rule
τ_0 - Initial Pheromone level
τ_s - Pheromone level in the tuple

Global Pheromone Trail Evaporation Rule: Applied at last after all the ants attain their migration plan. It is explained as

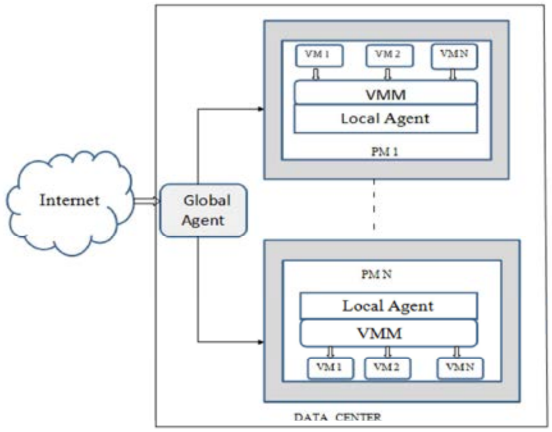

Fig. 2: System Model

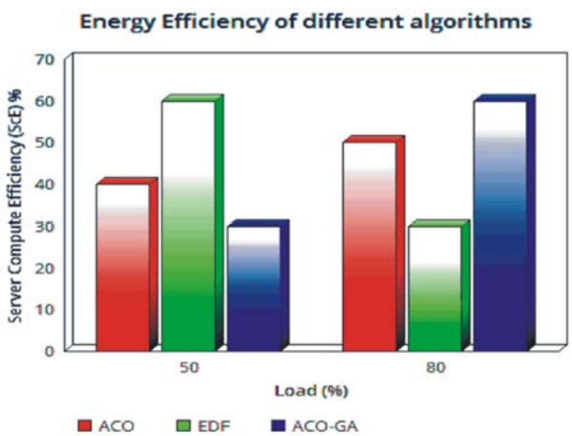

Fig. 3: Performance Measure

$$\tau_s = (1-\alpha).\tau_s + \alpha.\Delta_{\tau_s}{}^{+},$$

$$\Delta_{\tau_s}{}^{+} = \begin{cases} f(M^{+}), & \text{if } s \in M^{+} \\ 0, & \text{otherwise.} \end{cases}$$

α - Pheromone decay parameter in the global updating rule
Δ - Additional pheromone amount given only to the tuples belongs to the global best migration

The output of ACO algorithm is a set of tuples T where each tuple t ∈ T consists of three elements a source PM (*pso*), Destination PM (*pde*) and VMs to be migrated (v) and is represented by equation

t = (*pso*, v,*pde*)

GA algorithm (Mutation) is applied on the set T which will further refine the resulting tuples and will produce only the best set of tuples to be selected for VM migration.

System Model: Figure 2 represents the system model of the project. The proposed system model works on the basis of two types of agents : Local Agent and Global Agent.

Local Agent: Resides in a Physical Machine to solve the Physical machine's status detection by the usage of current resource utilizations of the PM.

Global Agent: Acts as a supervisor and optimizes the VM placement. Users can communicate with Global agent through internet. Efficient allocation of VMs to the PMs is performed using EDF/GA-ACO algorithms based on the threshold value.

Cloud Simulator: It is the tool used for Modelling and Simulation of cloud infrastructure and applications. Details such as number of data centres, number of physical machines and number of VMs are collected and simulated in cloudsim. Initially VMs are placed on PMs using MBFD algorithm. Then based on workload, algorithms are swapped between EDF and Hybrid ACO at regular intervals.

Result and Analysis: In this section, we compare proposed architecture which uses combination of EDF and hybrid ACO-GA algorithm with existing architecture which uses just ACO algorithm for dynamic allocation of VMs. The results of Figure 3 clearly shows, the server compute efficiency (ScE) of hybrid ACO-GA during over load and EDF during under load are higher than the traditional ACO algorithm. This indicates that usage our new architecture enables the data centre to achieve its maximum server compute efficiency in all the cases irrespective of the load.

Conclusion and Futurework: This paper focuses on the design and implementation of a distributed architecture, which is used for live VM migrations with the help of algorithms such as Modified best fit decreasing (MBFD),Earliest Dead line first (EDF) and the bio inspired dynamic hybrid ACO –GA. The main goal of this paper is to prove how the combination of algorithms used here can outperform ACO algorithm while used for live VM migration in terms of energy efficiency, limited SLA violation and time consumption. In the system proposed, checking the threshold to switch between the algorithms is done at regular intervals. As a future work, system can be designed in a way to automatically change the algorithm either to EDF or to hybrid ACO-GA as per the threshold variation.

REFERENCES

1. Bianchini, R. and R. Rajamony, 2004. Power and Energy Management for Server Systems, Computer, 37(11): 68-74.
2. Anshul Gandhi, Mor Harchol-Balter, Mike Kozuch and Ram Raghunathan, 2012. AutoScale: Dynamic, Robust Capacity Management for Multi-Tier Data Centers. ACM Transactions on Computer Systems, 30(4): 1-26.
3. Adnan Ashraf, Fahimeh Farahnakian, Hannu Tenhunen, Ivan Porres, Juha Plosila, Pasi Lilje berg and Tapio Pahikkala, 2015. Using Ant Colony System to Consolidate VMs for Green Cloud Computing IEEE Transactions on Services Computing, 8(2).
4. Clark, C., K. Fraser, S. Hand, J.G. Hansen, E. Jul, C. Limpach, I. Pratt and A. Warfield, 2005. Live migration of virtual machines, in Proc. 2nd Conf. Symp. Netw.Syst.Design Implementation.
5. Jian Cao, Minglu Li and Yihua Wu, 2012. Energy Efficient Allocation of Virtual Machines in Cloud Computing Environments Based on Demand Forecast, © Springer Science+Business Media, LLC 2012.
5. Cader, T., A. Rawson and J. Pfleuger, 2008. Green Grid Data Center Power Efficiency Metrics. Consortium Green Grid.
6. Young Choon Lee Albert Y. Zomaya, 2012. Energy efficient utilization of resources in cloud computing systems, Published online: 19 March 2010 © Springer Science+Business Media, LLC 2010.
7. Weijia Song, Qi Chen and Zhen Xiao, XXXX. Senior Member, IEEE, Dynamic Resource Allocation Using Virtual Machines for cloud computing Environment, IEEE Transactions on (Volume:24, Issue: 6)
8. Di Caro, M., G. Dorigo and L.M. Gambardella, 1999. Ant algorithms for discrete optimization, Artif. Life, 5(2): 137-172.
9. Dorigo, M. and L. Gambardella, 1997. Ant colony system: A cooperative learning approach to the travelling salesman problem, IEEE Trans. Evolutionary Comput., 1(1): 53-66.
10. Jashweeni Nandanwar and Urmila Shrawankar, 2013. Aggregation of EDF and ACO for Enhancing Real Time System Performance, International Journal of Computer Applications (0975 – 8887) Volume 73– No.17, July 2013
11. Irving VitraPaputungan andZainudinZukhri “a hybrid optimization algorithm based on genetic algorithm and ant colony optimization, International Journal of Artificial Intelligence & Applications 09/2013.