Academic Science

International Journal of Electronics, Electrical and Computational System
IJEECS
ISSN 2348-117X
Volume 6, Issue 8
August 2017

A Model for Energy Efficient Workflow Scheduling of Scientific Applications in Cloud

S. Balamurugan
Perunthalaivar Kamarajar Institute of Engineering and Technology (*PKIET*),Karaikal, India
S. Saraswathi
Pondicherry Engineering College (*PEC*),Puducherry, India

ABSTRACT

In recent times, there is rapid increase in the usage of cloud computing and its resources for the deployment of scientific workflows. All the scientific applications are large-scale in nature, so the successful execution of the application can be technically achieved by expanding in a cloud platform. Also, there is a growing demand for the cloud services and cloud providers not only in the field of deploying the scientific applications but also on the other application areas. A single scientific workflow usually contains thousands of tasks, requiring a large amount of computing resources for execution. These resources can be provisioned only by the cloud infrastructure. All the computing resources are provided in the form of Virtual Machines inside a cloud platform. A scientific workflow execution in cloud platform incurs huge energy consumption, so it is very important to deploy Virtual Machines in an energy efficient manner. So, energy consumption of a cloud computing resources has gained more attention among the researchers. It is still a challenging task to execute scientific workflow in an energy aware manner across the cloud platforms. So, in this paper, we propose a model for energy efficient workflow scheduling of scientific applications in cloud environment.

KEYWORDS

Workflow; energy-efficiency; scientific applications; resource provisioning; task assignment; task migration virutal machine; scheduling.

I. INTRODUCTION

Cloud computing is an internet based computing framework which provides large-scale processing elements which are distributed and located such that the resources and services are accessible to customers on demand. The important concept in cloud computing is virtualization. Virtualization is creating a virtual version of a device or a resource such as storage device, server, network or an operating system. Using a virtualization technology, we can able to convert a single physical server to run several virtual machines simultaneously. Workflow is defined as a group of tasks which are executed in a specific order to pertaining to any application. In recent days, usage of cloud computing and its resources for the execution of scientific workflows are gaining more attention.

Scientific workflows permit the users to easily define multi-step computational tasks, such as retrieving data from any product or a database, reorganize the data, analyze and run. Scientific workflows are accepted useful paradigm to describe, manage, and share complex scientific analyses. Scientific workflows have arisen to challenge the problem of excessive complexity in scientific experiments and applications. Different types of tasks that can be performed in a single workflow. Each task is responsible for a small piece of functionality such that many tasks need to be chained in a pipeline in an order to obtain a workflow that performs some useful task.

A Scientific workflow involves more than thousands of tasks and each task will have a sub-deadline and an overall deadline. However, with the large scale usage of data centers and computing servers, the energy consumption has become a big challenge for the cloud service providers. The US Environmental Protection Agency (EPA) has reported that in 2006 the IT infrastructures in USA have consumed more than 61 billion kWh for 4.5 billion dollars. And also the Energy utilization of a typical data center is getting increased for

around 20%-30%. This means that a pretty large amount of electricity energy is getting wasted in the usage of data centers and computing servers in the cloud environment. One effective way to reduce the power consumption is to consolidate the hosting workloads and shutdown physical machines which become idle after consolidation.

The remaining sections of paper are organized as follows. In Section 2, we have discussed about the related work and algorithms prevalent to energy efficient scheduling. Section 3 describes the energy models and its related equations for finding the energy consumption of a virtual machine, cloud application, and system used in this paper. In Section 4, we introduce our proposed model for Energy Efficient Workflow Scheduling of Scientific applications and finally, Section 5 concludes the paper with future work.

II. RELATED WORKS

In this section we will discuss about the various energy efficient workflow scheduling algorithms presented by many researchers.

Xiaolong Xu et.al. Proposed an Energy-aware Resource Allocation called EnReal [1], to address the problem of energy consumption of workflow tasks. It influences the dynamic deployment of virtual machines for scientific workflow executions. In this work, when deploying Virtual Machines for workflow task execution, it mainly targets on reducing the energy consumption which includes the Physical Machine baseline energy consumption, the energy consumed by idle VMs, idle PMs and the dynamic operations. To achieve the maximum energy savings all the idle PMs are switched from high to low power mode or it is put into sleep mode dynamically. The total energy consumption E can be calculated by

$$E = E^{AppExe} + E^{DynaOp},$$

Where, E^{AppExe} denotes the energy consumption for application execution and E^{DynaOp} denotes the energy consumption for dynamic operations.

Zhuo Tang et.al. Proposed an Energy-Efficient Task Scheduling Algorithm [2] in DVFS-enabled Cloud Environment. To achieve energy efficiency, this author proposed a DVFS-enabled Energy-Efficient Workflow Task Scheduling algorithm called DEWTS. In this algorithm, the relatively inefficient processors are merged by reclaiming the slack time. After the several recurrent merges it can leverage the useful slack time. DEWTS uses HEFT algorithm for initial scheduling of tasks. After that the underutilized processors are merged by closing the last node and redistributing the assigned task on it. During the task slacking phase, the tasks are distributed in the idle slots under lower voltage and frequency using DVFS technique. The experimental results showed that DEWTS has reduced the total power consumption by up to 46.5% for various applications.

Huangke Chen et.al. developed energy efficient online scheduling Algorithm called ENOS [3] for real time workflows. In order to improve the energy efficiency the author in this paper proposed three strategies for scaling up and down the computing resources and integrated into ENOS to balance weighted square frequencies of hosts.

Khadija Bousselmi et.al. Proposed an energy aware scheduling method for scientific workflows using Workflow Partitioning for Energy Minimization (WPEM) [4] algorithm and Cat Swarm Optimization (CSO). The algorithm has been devised in two stages; in first stage they proposed a Workflow Partitioning for Energy Minimization (WPEM) algorithm which reduces the energy consumption of the workflow and the total amount of data communication while achieving a high degree of parallelism. In the second stage, they have used the heuristic of Cat Swarm Optimization (CSO) to schedule the generated partitions so as to minimize the overall energy consumption of the workflow and its execution time.

The energy consumed for the execution of entire workflow is given by the equation

$$P_w = \sum_{(i=1)}^{n} \sum_{(j=1)}^{m} x_{ij} * (P_{ij} + \sum_{(k=i+1)}^{n} y_{ij} * Dt_{out_i} * E_{ik}/d_{ik})$$

Where, n is the total no. of tasks in the workflow, m is the total no. of virtual machine, Dt_{out} represents the amounts of data generated after the completion of the workflow task. The proposed algorithm has been evaluated using three real applications of data intensive workflows and it is compared it with other algorithms from literature. The experimental results show that their proposed algorithm has remarkable reduction of energy consumption with the tested workflows. The energy savings of up to 96% of network energy consumption for memory intensive workflows and the overall energy consumption of the workflows with a reasonable execution time and using less Cloud resources usage.

In [5] authors presented an energy efficient task scheduling algorithm for scientific workflows using frequency scaling and strictly following user deadline. Their work focused on heterogeneous machines with different task runtime and frequency capabilities. Initially tasks of the workflows re mapped to available machines. Energy Aware Stepwise Frequency Scaling (ESFS) algorithm determines the processor frequency to be used for the execution of each task taking into account its deadline and ensure that deadline was not violated. The proposed algorithm works iteratively to gradually scale the processor frequency used for each task for as long as overall energy savings are increased. The next available frequency mode for each processor is used as a lower bound in each iteration to scale the frequency of the tasks. ESFS algorithm is implemented on LIGO, SPIHT and Montage workflows and results showed that energy consumed is reduced when compared with baseline algorithms HEFT and EES.

In [6] the author presented a new Pareto-based multi-objective workflow scheduling algorithm. It is the extension of an existing state-of-the-art heuristic capable of computing a set of tradeoff optimal solutions in terms of makespan and energy efficiency. This approach is based on empirical models which capture the real behavior of energy consumption in heterogeneous parallel systems.

III. ENERGY MODEL OF CLOUD

Normally cloud will offer different types of Virtual Machine (VM) like $v_1, v_2, v_3 \ldots v_k$. All the VMs have different computing performance and used for different type of tasks. There are many characteristics of Virtual Machine e.g. computing performance C calculated in million instructions per second, maximum power P and bandwidth B. There is a strong relation between computing performance of a virtual machine and power. If a virtual machine of any type has higher computing performance then it consumes more CPU power.

$$\text{If } C_k > C_{k+1} \text{ then } P_k > P_{k+1} \quad (1)$$

The energy consumption of any virtual machine is of two type's static and dynamic energy consumption. In this paper we only consider the dynamic energy consumption and the energy spend by the virtual machine in idle state. Let E_d denotes the dynamic energy consumption and $E_{i\iota}$ denotes the idle state energy consumption. P_k denotes the power consumption of any virtual machine v_k. It is calculated by the equation

$$P_k = V_k^2 * f_k \quad (2)$$

Here, V_k is the voltage level of virtual machine v_k and f_k is the execution frequency v_k.

Now, the energy consumption can be defined as the product of power consumption of virtual machine and time. So, the dynamic energy consumption in a given time t can be calculated by the equation

$$E_{dy} = P_k * t_e \quad (3)$$

The static energy consumption can be calculated by the equation

$$E_{i\iota} = P_k * t_{i\iota} \quad (4)$$

So, the total energy consumed by any virtual machine is calculated by the equation

$$E(v_k) = E_d + E_{i\iota} \quad (5)$$

When the jobs are submitted to the cloud, the physical machine provides the hardware infrastructure. Based on the task request Virtual Machines are created and the tasks are executed on the virtual machine.

International Journal of Electronics, Electrical and Computational System
IJEECS
ISSN 2348-117X
Volume 6, Issue 8
August 2017

IV. PROPOSED MODEL

Before proceeding with the algorithm, we introduce a concept called power utility. Power Utility is defined as the ratio of workload to the total energy consumption of task t_k on virtual machinevm_k. Workflow Task is mapped to any virtual machine based on the Power Utility [7] concept.

The Power Utility is given by the equation

$$\text{PU}\ (t_k) = \frac{W_i}{E\,(v\ \ _k)} \qquad (6)$$

Now, we will form a Power Utility Set called PU (t_k) for task t_k which includes all the power utilities when that task is executed in different various virtual machines. Likewise, for all the tasks present in the workflow, a power utility set is formed. The set is given as

$$\text{PU}\ (t_k) = \{\text{PU}\ (t_1),\ \text{PU}\ (t_2)\ldots\ \text{PU}\ (t_k)\}. \qquad (7)$$

The power utilities in the set are arranged in descending order. The first Power Utility denotes the optimal power utility for executing the task t_k on the virtual machine and the corresponding virtual machine is denoted as $v\ \ _o$.

Fig.1 illustrates the working of task assignment model. Let us assume ‘n’ no. of virtual machines and ‘n’ no. of tasks extracted from the workflow DAG. A set of Tasks and virtual Machines are given to the power utilization module. Each Virtual Machine will have different power utility for each task. In the power utilization module, the power utility is calculated for each virtual machine. Now, a power utility set is formed for each task in the task list. The Power Utility is given by the equation 6. Based on the PU list, each task is assigned to an optimal virtual machine. This is done is task assignment block. The entire power utility list is arranged in descending order. The first Power Utility denotes the optimal power utility for executing the task t_k on the virtual machine and the corresponding virtual machine is denoted as vm_o .

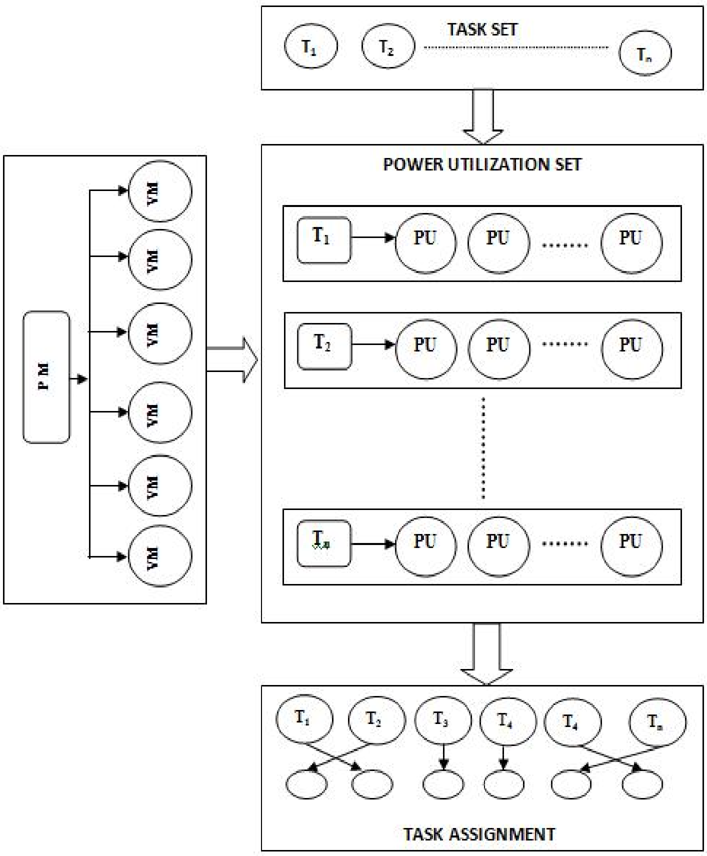

Fig. 1: Task Assignment in EAWSTM

Academic Science

International Journal of Electronics, Electrical and Computational System
IJEECS
ISSN 2348-117X
Volume 6, Issue 8
August 2017

In this method, finding the optimal virtual machine for the execution of task is done prior to assigning the task. Prior estimation is done on the virtual machine based on the nature of the incoming task, specification of the virtual machine and historical data. So, this method can guarantee that the given task will be executed on the optimal virtual machine, so that the overall energy efficiency can be improved.

Next, if two tasks t_i and t_{i+1} have already been mapped to Virtual machine of type vm_i and $v\ \ _{i+1}$. If it is found that task t_{i+1} is the successor of task t_i. The energy consumption of the two tasks are E (t_i) and E (t_{i+1}). If these two tasks are migrated to $v\ \ _k$ the execution time of two tasks can be computing by the formula

$$T\ (t_i, t_{i+1}) = W_i + W_{i+1} / C_k + D_T \qquad (8)$$

Here, D_T is the data transfer time, but we will need the data transfer time the two tasks are going to be migrated on the same virtual machine $v\ \ _k$.

This migration can be done only if it satisfies the following condition

$$\text{deadline}\ (t_i) <= \text{deadline}(t_i) + \text{deadline}(t_k) \qquad (9)$$

If the above condition is not satisfied then we will not migrate the task and remove it from the TASK set. Repeated the above step till the TASK set is empty.

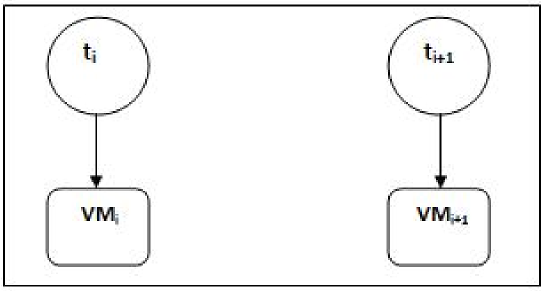

Fig.2 Individual Task running on VM's

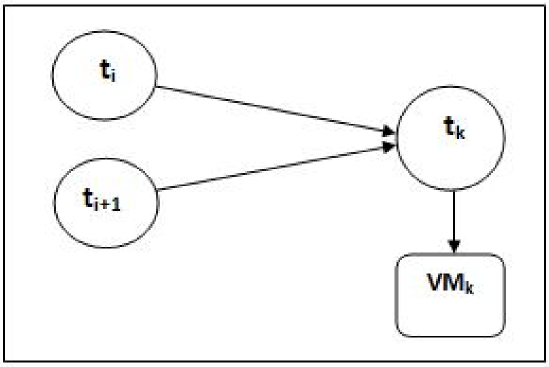

Fig.3 Task Merging and Migration

Once the task assignment is done successfully, then we need to find whether there is any possibility for task migration. Fig. 2 shows the task assignment for task t_i task t_{i+1} on VM_i and VM_{i+1} respectively. Fig.3 shows the migration of two tasks based on the deadline. In the task set find if task t_k can be finished within the deadline of task t_i, create a virtual machine $v\ \ _k$ and migrate the two task to $v\ \ _k$. The migration should takes place based on the condition said in equation 9. If it is possible for us to merge and migrate more no. of tasks onto a fewer virtual machines, we can able switch off one or more physical machines which does not have any virtual machines. So, it is possible to conserve more energy when we execute the workflow on cloud.

V. CONCLUSION AND FUTURE WORK

In this paper we consider the problem of higher energy consumption for the execution of workflow in cloud virtual machines. The related works pertaining to energy efficient workflow scheduling are analyzed.

We introduced our model for energy Efficient Workflow Scheduling of scientific applications. This model is composed of two stages called task assignment stage and task migration stage. After the completion of these two stages, we will get the better allocation scheme i.e. workflow task scheduling policy where all the task are mapped to optimal virtual machines for execution, through which we can attain better energy consumption.

As a future work, we plan to evaluate the proposed algorithm using different scientific workflows and subsequently access its energy consumption when scheduled to virtual machines.

REFERENCES

[1] Xiaolong Xu, Wanchun Dou, Xuyun Zhang, and Jinjun Chen, “EnReal: An Energy-Aware Resource Allocation Method for Scientific Workflow Executions in Cloud Environment” IEEE Transactions on Cloud Computing, Vol.4, No.2, pp.166-179, April-June 2016.

[2] Zhuo Tang, Ling Qi, Zhenzhen Cheng, Kenli Li , Samee U. Khan, Keqin Li, “An Energy-Efficient Task Scheduling Algorithm in DVFS-enabled Cloud Environment” Journal of Grid Computing, Springer Science, DOI 10.1007/s10723-015-9334-y, March 2015.

[3] Huangke Chen, Xiaomin Zhu*, Dishan Qiu, HuiGuo, Laurence T. Yang, peizhong lu, EONS: Minimizing Energy Consumption for Executing Real-Time Workflows in Virtualized Cloud Data Centers”, IEEE International Conference on Parallel Processing Workshops, DOI 10.1109/ICPPW.2016.60, 2016.

[4] Khadija Bousselmi, Zaki Brahmi, Mohamed Mohsen Gammoudi, “Energy efficient partitioning and scheduling approach for Scientific Workflows in the Cloud” IEEE International Conference on Services Computing, DOI 10.1109/SCC.2016.26, 2016.

[5] Pietri, Ilia, and Rizos Sakellariou. "Energy-aware workflow scheduling using frequency scaling." In Parallel Processing Workshops (ICCPW), 2014 43rd International Conference on, pp. 104-113. IEEE, 2014.

[6] Juan J. Durillo, Vlad Nae, Radu Prodan, “Multi-objective energy-efficient workflow scheduling using list-based heuristics”, Future Generation computer systems, Vol.36, pp.221-236, 2014.

[7] Hao Li, Hai Zhu, Guoheng Ren, Hongfeng Wang, Hong Zhang, Liyong Chen, “Energy Aware Scheduling of Workflow in Cloud Center with Deadline Constraint”, 12th International conference on Computational Intelligence and Security, IEEE, 2016.