TEST Engineering & Management

January - February 2020
ISSN: 0193 - 4120 Page No. 6056 - 6070

Scheduling Workflows with Reduced Energy Consumption for Big Data Applications

V. Seethalakshmi
Research Scholar Dept. of Computer Science and Engineering Pondicherry Engineering CollegeSeethalakshmi.v@pec.edu
Dr. V. Govindasamy
Associate Professor Dept. of Information Technology Pondicherry Engineering College vgopu@pec.edu
Dr. V. Akila
Assistant Professor Dept. of Computer Science and Engineering Pondicherry Engineering College akila @pec.edu

Article Info
Volume 82
Page Number: 6056 - 6070
Publication Issue:
January-February 2020

Article History
Article Received*: 18 May 2019*
Revised*: 14 July 2019*
Accepted*: 22 December 2019*
Publication*: 29 January 2020*

Abstract:
Abstract—The consumption of energy and the cost have a great impact in influencing the consumption of resource and economy. Dueto this fact, the two pivotal factors-the consumption of energy and the cost- play a vital role in contributing to the execution of application in cloud computing systems. This article highlights various issues of minimizing the consumption of energy of a cost budgeted Directed Acyclic Graph (DAG) that can be applied in heterogeneous computing systems. In this paper, cost and energy aware scheduling is proposed using Reduced Energy Consumption using the Available Budget Pre-assignment (RECABP) technique for the cloud scheduler in the reduction of the execution cost of workflow.RECABP also enables in reducing the energy consumption. In addition to the minimization of the execution of workflow cost and the consumption of energy, the time complexity analysis is performed to ascertain that the complexity of RECABP algorithm is polynomial. The proposed RECABP algorithm is implemented using CloudSim. The implementation demonstrates that the proposed RECABP algorithm outspaces the related well-known approaches in terms of the following parameters - computation cost, execution time, bandwidth preference, time preference and energy consumption.

Keywords: *Cloud Computing, Scheduling Algorithm, Workflow, Big Data*

I. INTRODUCTION

Big Data refers to the huge volume of both organized and unorganized information that has been created on a regular basis. The major issues in Big Data are how it can be managed efficiently and theimmense data is analyzed.The availability of cloud computing is on-demand. It cannot be achieved, without direct active management by the user. Big Data and Cloud Computing attract global attention because of their multiple business-driven promises and desires, lower upfront IT costs, quick time to market the opportunities for making inventive business. Distributed computing guarantees a scalable infrastructure of large amount of information on Cyber-Physical Environment [1].On combining Big Data with powerful analytics, we can achieve business related jobs. The data centers processthe large clusters amidst using the application of Big Data. It has been a

January - February 2020
ISSN: 0193 - 4120 Page No. 6056 - 6070

critical concern on executing the applications of energy costs. Big Data is of use in the diversified fields, including health care, financial service and commercial recommendation.

The Economistquotes that Big Data is emerging as a new raw material of business. Nowadays, the information to be examined is dynamic and enormous in volume. In addition, it is of multiple information types. This information originates from various information sources, for example, Facebook, Whatsapp, Twitter, YouTube, GPS signals of mobile phones and more. Hence, Big Data possess the different attributes such as heterogeneity, unstructuredness, semi structuredness, incompleteness, high dimensionality.

Distributed computing domain has proposed several job scheduling algorithms. With the help of cloud environment, most of them can be applied with suitable modifications. The main objectives of job scheduling are to achieve a better performance evaluation and the good throughput [2]. In cloud environment, the traditional job scheduling algorithms are not able to provide scheduling.

1.1 Scheduling Workflow Applications in Cloud

(i) Workflows

The application of workflow is generally a specific outcome which is executed in a sequence that are considered to be a collection of multiple job. These jobs are executed depending on their information dependence. The relationship between these jobs is related to the relationship of parents and children. The child's task is only performed after the parent's task is performed. Directed Acyclic Graph (DAG) is applied in any application in workflow. DAG is represented as G (v, e) wherein Graph is 'G'. Here, 'v' indicates the number of nodes and data between jobs represented by 'e' with respect to data dependencies [3]. The Figure-1 indicates a workflow.

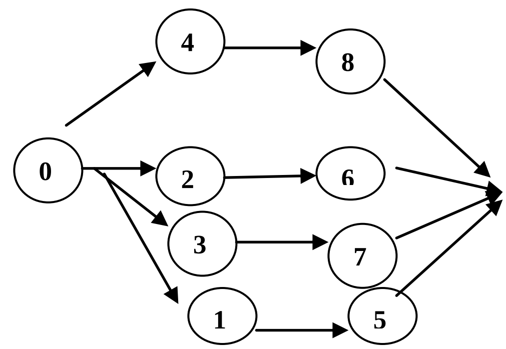


Figure 1. Workflow Representation in Graph

Figure 1 displays independency in a graph G among various jobs. Jobs 1 to 4 performed after parent task 0 as they are child jobs. The output of parent node is given as input to the child node. The entry node task acts in 0 and the exit node task acts in 9. Task 9 is performed after completion of jobs 5, 6, 7 and 8. The entry task is not required for the parents. The exit task is not mentioned for any child. The main consideration of task scheduling techniques is only one entry and exit task. Task 0 acts as an entry node and task 9 acts as an exit node.

Makespan is featured by performance workflow. TheMakespan is computed as the difference between finishing time and beginning time of a workflow. Each edge and node is denoted graphically in the form of specific weight. The values of the nodes and edges of DAG are noticeable when separate methods are applied in computing.

(ii) Scheduling in Cloud Computing

Distributing resources to the specific job in a particular time is known as scheduling. The effective purpose of scheduling is to make full use of the resource.The resource allocation for workflow applications is quite challenging in the cloud. Although many heuristic algorithms for scheduling have been suggested, improvements are required to make the machine speed and more active. The original algorithms for scheduling include First Come First Serve, Shortest Job First, Round Robin, Min – Min and Max - Min. This scheduling technique definitely does not yield a best outcome with cloud computing for scheduling problems [4].

(iii) Types of Scheduling

Figure 2 highlights the types of scheduling algorithms. Static Scheduling:All information about the jobs and the resources are known to the schedulerbefore execution.When an overhead is provided less runtime is utilized in the case of static scheduling.

Dynamic Scheduling:Here, the data on the job components is unknown before execution. In addition, Job Arrival time is also unknown.In this type of scheduling runtime overhead is greater.

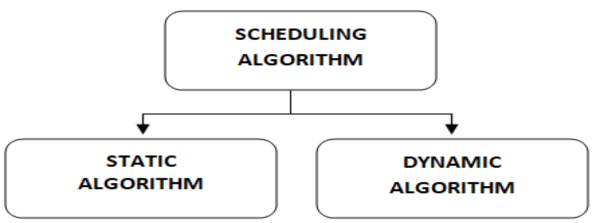


Figure 2. Types of Scheduling

(iv) Static Scheduling Algorithms

Before running any workflow jobs, the static scheduling technique is based on creating a schedule plan. That is, it considers the cloud platform's initial status. These algorithms can be used to schedule one or more workflows.

(v) Dynamic Scheduling Algorithms

Dynamic algorithms address the provision of resources and the scheduling of runtime jobs to take into account the cloud's dynamic aspect. The mapping solutions are recalculated periodically to optimize computational costs based on resources and current network.

II. RELATED WORK

Overview of Strategies to Schedule Workflow Applications in Cloud

Minimizing energy consumption for workflow applications is the important challenge in the cloud. The execution time of the user interface solely relies the amount of resources allocated and how the financial cost is being determined. Many researchers have found new ways to overcome the challenges. Several algorithms have been proposed to schedule Workflows.

(i) Multi-Objective Privacy-Aware (MOPA) Workflow Scheduling Algorithm:

MOPAalgorithm helps in providing the cloud customer with a set of Pareto tradeoff solutions [5]. The problem-specific encoding and population initialization are proposed in MOPA algorithm. The quality solutions obtained by this algorithm generally outperform the algorithms of Modified Non-dominated Sorting Genetic Algorithm-II (MNSGA-II) and Modified Multi-objective Particle Swarm Optimization (MMOPSO) with the growth of the size of workflow instances.

(ii) Modified Particle Swarm Optimization (MPSO) Scheduling Algorithm:

The aforesaid MPSO algorithm outperforms the other state-of-the-art systems as far as normal average response time, execution time and maintaining varied instances of balancing the load of the Virtual Machines (VM). The main ideology of the proposed algorithm is to utilize cloud resources effectively and also to improve execution cost of batch of task in the heterogeneous cloud environment based on IaaS [6].

(iii) Partition Problem Based Dynamic Provisioning and Scheduling (PPDPS) Algorithm:

PPDPS is used to cut-down the cost of execution of scheduling a time-constrained workflow. The

January - February 2020
ISSN: 0193 - 4120 Page No. 6056 - 6070

combined nature of k- means clustering technique and partition problem is most importantly proposed for the dynamic provisioning [7]. Analysis of Variance (ANOVA) test is performed. Its post-hoc analysis is done utilizing the Least Significance Difference (LSD) test to ascertain the better performance of the PPDPS algorithm.

(iv)Completion Time Driven Hyper-Heuristic (CTDHH):

Dynamic Hyper-Heuristic Algorithm (DHHA) is the major role of the CTDHH proposed method. It works on a huge search space instead of finding solutions independently by implementingthe components of heuristic or meta-heuristic of set [8]. The performance is aggregated by differentiatingit with four population-based methods and also with a previous hyper-heuristic method called the Hyper- Heuristic Scheduling Algorithm (HHSA).

(v) A Novel Three-Phase Approach:

This approach includes the initial algorithm for scheduling, the task removal algorithm, and the Dynamic Voltage and Frequency Scaling (DVFS) algorithm to aware reliability. Cloud workflows produce energy-efficient job levels under both reliable condition and average time limitations [9]. Execution and results from variant traditional workflow milestones demonstrate the efficiency of the Novel Three- Phase method. The Novel Approach achieves more than 30 percent energy with no infringing reliable condition and execution time limitations.

(vi) Hybrid Algorithm (Ant Colony Optimization and Cuckoo Search):

In order to minimize energy consumption, the voltage scaling factor is focused by hybrid algorithm. The proposed algorithm initializes the initial solution of pheromone, heuristic information, nest number and random solution [10]. The energy used by the hybrid algorithm decreases as the number of jobs increases considerably and energy consumption remains in the steady state at the particular number of jobs. The consumption of energy is calculated as well as the rate of improvement is compared with an existing Ant Colony Optimization (ACO) algorithm for a number of jobs.

(vii) Dynamic Energy Aware Scheduling Algorithm:

This algorithm attempts to use the parameter of energy use and the time of execution of jobs depending on rank model.The consumption of energy and the time of execution are considered in the Dynamic Energy Aware Scheduling Algorithm. The scheduling method describes about transfer rate and its related task distance to assign jobs a multi-dimensional scheduler framework is formulated [11]. To prevent many specific task transfers, the scheduler keeps the related jobs in the nearest distance. Additionally, estimation of task behavior is managed by VMs are managed by estimating the task behavior to avoid shutting down and starting unnecessary VMs.

(viii) Evolutionary Multi-Objective Optimization (EMO):

EMO minimizes both makespan and cost simultaneously on an IaaS platform. Instead of simulating the cloud into a traditional heterogeneous service pool, the encoding scheme and genetic operators are designed directly based on the IaaS model to reduce the search space, to improve search capacity and to simultaneously accelerate search speed [12]. The proposed algorithm is compared with the various scheduling algorithms includingMulti-Objective Heterogeneous Earliest Finish Time (MOHEFT),Non-dominated Sorting Particle Swarm Optimizer (NSPSO), εFuzzy Particle Swarm Optimization (PSO), Strength Pareto Evolutionary Algorithm (SPEA*) and Multi-Objective Differential Evolution (MODE). Experimental outputs evinces that the EMO algorithm is more stable and produces acceptable schedules.

(ix) Cost and Energy Aware Scheduling Algorithm (CEAS):

January - February 2020
ISSN: 0193 - 4120 Page No. 6056 - 6070

The proposed module enables cloud schedulers to spend minimal costs on completing a workflow as well as minimizing energy consumption without exceeding the time limit [6]. Comparative assessment is done on the basis of energy consumption with two well-known scheduling algorithms, Energy-Efficient Scheduling algorithm (EES) and Heterogeneous-Earliest-Finish-Time (HEFT) technique. Each sub-algorithm's time complexity is polynomial so that the monetary cost can be reduced.

(x) DVFS-Enabled Energy-Efficient Workflow Task Scheduling Algorithm: DEWTS

DEWTS the initial scheduling order jobs producing the entire makepan and deadline.The most important achievement to minimize the processing energy costs. To examine the overall performance of the DEWTS algorithm, the differentiation is made with two heuristic algorithms: HEFT and EES [9]. Results show that the DEWTS algorithm can be applied to heterogeneous cloud environment to different parallel applications. This algorithm significantly reduces energy. In addition, Quality of Service is achieved by complying with pre-set deadline.

III. Analysis of Scheduling Algorithms

Table 1.Analysis of Scheduling Algorithms

	Performance Metrics								
	Decrease of Execution Time	DecreaseofMakespan	Improvement Resource Utilization	Improvement CPU utilization	Deadline Constraint	Improvement QOS	Decrease of Cost	Decreaseof Time complexity	Decreaseof Energy Consumption
Round Robin and First-Fit Algorithm									
Multi-Objective Privacy Aware Algorithm									
Modified Particle Swarm Optimization (MPSO) Scheduling									
Partition Problem Based Dynamic Provisioning and Scheduling									
Completion Time Driven Hyper-Heuristic approach									
Reliability-Aware DVFS Algorithm									
Hybrid Algorithm (Ant Colony and Cuckoo Search)									
Dynamic Energy Aware Scheduling Algorithm									
Evolutionary Multi-objective Optimization									
Cost and Energy Aware Scheduling Technique									
DVFS-Enabled Energy-Efficient Workflow Task Scheduling	✓					✓		✓	✓

January - February 2020
ISSN: 0193 - 4120 Page No. 6056 - 6070

In this table, comparative analysis of Scheduling algorithms is done by comparing the various performance metrics. It shows the best algorithms in respect to its Performance measure. For Eg: Round Robin and First-Fit Algorithm shows the best performance in Reducing Execution time, makespan and Energy Consumption. Variant performance metrics are being used to calculate the execution of algorithms

IV. EXISTING SYSTEM

The existing system provides analytical method to reduce the entire amount of consumption of energy and to analyze the efficiency of energy for a scheduling algorithm.

For evaluating total energy consumption, the analytical methods use the following two scheduling methods:

(i). Always Switch-on Physical Machines (PMs) once put-on and

(ii). Put-off (hibernating) muted PMs during calculation, with two sufficient VM migrations.

Using Lower Bound-Min-Migration under Strongly Divisible Capacity Configuration, the existing system offers analytical solutions for the above two open methods, The first method is for Lower bound of final consumption of energy and another is for calculating energy-efficient of a given techniques.

4.1 The Lower Bound of the Energy Consumption

The lowerlimit of consumption of energy is determined in an open data collection by job conf iguration (VM requests) and PMs. For MethodA,whole energy use is disturbed by the total switch-on time of all PMs for the available group of VM jobs..

For Method B, once a group of VM tasks are provided, the final count of PMs (m) and their final switch on time (T).This helps in reducing the power consumption time.

4.2 Issues in Existing System & Inference

A scheduling mapping arranges the execution of a job on the availablesources of interdependent jobs. Scheduling allocates the appropriate resources to workflow jobs to complete the execution in order to fulfill the objective user-imposed functions. Proper scheduling has a major impact on system performance. NP-hard problems belong to the issues of mapping jobs on distributed services. No known algorithms can produce the best result for such problemsin polynomial time.The result depends on intensive search square measure impractical because the overhead is extremely high forgenerating schedules.

There are two methodologies of flow of work ordering. They are Best-effort-based scheduling and QoS constraint-based scheduling. The best-effort based scheduling reduces the execution time by take no notice of other factors such as the economic cost of accessing the resources and the QoS satisfaction levels of various users.

The QoSbased scheduling limitation minimizes efficiency under most significantQoSlimitations, such as time reduction under budget constraints or cost minimization under time limitations.

• The total amount of consumption of energy is further reduced
by limiting the amount of VM migrations,

• The Process execution time is more in LB-Min-Migration Algorithm.

• LB-Min-Migration Algorithm creates security and privacy issues and complexity level is very high.

• LB-Min-Migration Algorithm is not robust and flexible to execute their task in energy consumption budget mechanism.

• To minimize the amount of VM migrations, the LB-Min-Migration technique is essential as regular VM migration can cause congestion and vibration in the network.

January - February 2020
ISSN: 0193 - 4120 Page No. 6056 - 6070

•Other overheads such as degradation, overloaded PMs and management costs during VM migrations should also be considered.

V. Directed Acyclic Graph Model:

A cloud computing environment consists of a graph of heterogeneous Virtual Machine tosupply with various costs.

$U= \{u_1, u_2, ..., u_{|u|}\}$

This notation denotes a group of different virtual machines size, that |U| represents the countof thegroup U.

A process being executed is denoted by a DAG, G=(N, W, M, C)

(1) N denotes group of nodes in G, and a task is emphasized by each node n_i € N. Pred (n_i) refers to the set of n_i immediate predecessor jobs. Success (n_i) emphasizes the set of n_i immediate successor jobs. The n entry task is the task that has no predecessor and n exit task is the task that has no successor. Suppose a DAG program is having different go in or exitjobs, the graph will be accompanied by a duplicate go in or exitjobs with 0 - weight non-independents.

(2) W is a matrix of |N| * |U|, where N (set of nodes comprising of multiple jobs in virtual machine) * U (set of sizes of virtual machine). Each entry,$W_{i,k}$denotesthe time of execution of$n_i$executing on u_k. Due to the heterogeneity of VM, for single task n_i is having variantarrival time scores on the variant VMs.

(3) Mis a communication edge set and each edge $m_{i,j}$ € M representsthe communication from n_i to n_j.

(4) $c_{i,j}$€ C refers to the communication time of $m_{i,j}$whenn$_i$ and n_j are not allocated to the equal virtual machine. The count of information transmittedamong the jobn$_i$and the task $n_j$is based on the Communication time $c_{i,j.}$ This is not related to the computation service of the same virtual machine. $c_{i,j}$ is zero when both the jobs are attached to the same VM.

Figure.3 depictsaninspiringillustration of a DAG application. Table 1 demonstrates a time value matrix in Figure. 3. The example illustrates 10 tasks executed on 3 VMs $\{u_1, u_2, u_3\}$. The weight 16 of n_1 and u_2 in Table 1 represents the execution time with the maximum frequency denoted by $w_{1,2}$=16. The weight 18 of the edge between n_1 and $n_2$depicts the communication time depicts as $c_{1,2}$if n_1 and n_2 are not assigned to the same VM.Every time unit is exclude in this example for simplicity.

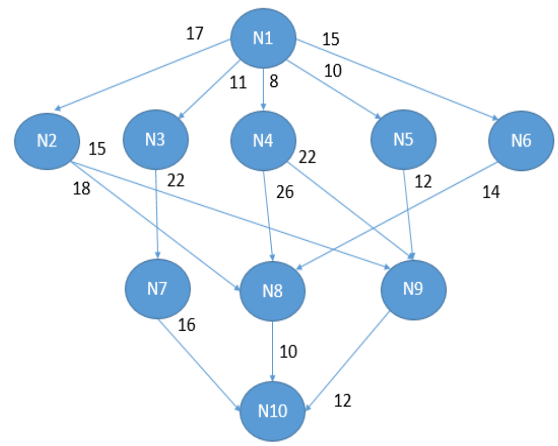


Figure 3. Example of a DAG application with ten tasks

Table1.implies the time of execution values of tasks on varied VMs with the maximum frequencies of the DAG application in Figure. 3.

Table 1:Execution time values of tasks on different VMs

Task	u_1	u_2	u_3
N_1	15	15	8
N_2	12	18	17
N_3	10	12	18

January - February 2020
ISSN: 0193 - 4120 Page No. 6056 - 6070

N_4	12	7	16
N_5	11	12	9
N_6	12	15	8
N_7	6	14	10
N_8	4	10	13
N_9	17	11	19
N_{10}	20	6	15

5.1Task Prioritization:

RECABP uses the task priority's upward rank value ($rank_u$). The $rank_u$is computed based upon the nodes assigned to the task. The more number of nodes assigned to the task, the $rank_u$ will be higher. In the beginning, all jobs of the DAG application are unassigned. Tasks are designated according to the degradation of upward rank value ($rank_u$). Thejobs are ordered depending of the descending $rank_u$order. The task assigned is $n_{s(y)}$, where $n_{s(y)}$ denotes the y^{th}allocated task, then $\{n_{s(1)}, n_{s(2)},\ldots, n_{s(y-1)}\}$ denotes the task set where the jobs are assignedand $\{n_{s(y+1)}, n_{s(y+2)},\ldots, n_{s(|N|)}\}$ denotes the task set where the jobs are unassigned. Initially, all DAG application jobs are not assigned and the jobs are prioritized based on the decreasing rank value ($rank_u$) order.

$$rank_u(n_i) = w_i + \max n_j \in succ\,(n_i)\ \{C_{i,j} + rank_u(n_j$$

5.2Energy Model

There is a direct connection between frequency and voltage. Dynamic Voltage and Frequency Scaling (DVFS)ismainly effort to save energy reducingthe both frequency and voltage. The changingvoltage and frequency are referred by the term frequency change simultaneously. The system-level power is also adopted for the DVFS-capable system. Therefore, the consumption of powerat frequencyP(f) of a frequency 'f' computing system is calculated as follows:

$$P(f) = P_s + h(P_{ind} + P_d) = P_s + h\,(P_{ind} + C_{ef}f^m)\ldots\ldots.(2)$$

Where

i. P_s denotes the static power which can only be detached if the complete system is changed.
ii. P_{ind} denotesthe dynamic power of frequency independent which may be deleted along withswitching machine into the non-working state.
iii. P_d denotes frequency-dependent active power.
iv. Hdenotesthe machine condition and represents whether active power is presently acceptingin the system. If the machine is active, h. =1; otherwise, h. =0.
v. C_{ef}denotes a capable switching capacity, and
vi. Mdenotes the dynamic power index that is not less than 2.

Both C_{ef} and mare VM- are exiting not independent constant.

In order to avoid excessive overhead, the system is always turned on. In other words, $P_s$ingests and is manipulated. This paperthrows light on active power (e.g., P_{ind} and P_d) and ignores the P_s in the computation. WhenP_d is less, the outcome in mean thatminimal is consumed because of the P_{ind},that is a least energy-efficient frequency fee existsand is denoted as

:

$$f_{ee}=\sqrt{m\sqrt{\frac{P_{ind}}{(m-1)C_{ef}}}}$$

Assuming the frequency of a VM varies from a leastexisting frequency f_{min} to the atmost frequency f_{max}, the lowest frequency to execute a task iscomputed as

$$f_{low} = \max(f_{min}, f_{ee})$$

Therefore, any actual valuable frequency $f_h$should belong to the scope of$f_{low} \leq f_h \leq f_{max}$.

TEST Engineering & Management
January - February 2020
ISSN: 0193 - 4120 Page No. 6056 - 6070

To contemplate that the number of VMs is |U| in the system and these VMs are completely heterogeneous, each VM should have separate power parameters.

Here,

1. We define frequency independent dynamic power set containing ‘P_{ind}’and its range starts from 1 to |U|. $\{P_{1,ind}, P_{2,ind},\ldots, P_{|U|,ind}\}$

2. It also considers frequency- dependent dynamic power set contains ‘P_d’ and its range starts from 1 to |U|.$\{P_{1,d},P_{2,d},\ldots, P_{|U|,d}\}$

3. It adds valuable switching capacitance set contains ‘C_{ef}’ and its range starts from 1 to |U|.
   a. $\{C_{1,ef},C_{2,ef},\ldots,C_{|U|,ef}\}$

4. With dynamic power exponent set contains ‘m’ and its range starts from 1 to |U|.
   a. $\{m_1,m_2,\ldots,m_{|U|}\}$

5. The least energy-efficient frequency set contains ‘f’ and its range starts from 1 to |U|.
   a. $\{f_{1,ee}, f_{2,ee},\ldots, f_{|U|,ee}\}$;
6. And at the last original effective frequency set contains ‘f’ and its series with variables low, max α, β and so on.

$$\{f_{1,low}, f_{1,\alpha}, f_{1,\beta}, \ldots, f_{1,max}\};$$
$$\{f_{2,low}, f_{2,\alpha}, f_{2,\beta},\ldots, f_{2,max}\};$$
$$\ldots;$$
$$\{f_{|U|,low}, f_{|U|,\alpha}, f_{|U|,\beta}, \ldots, f_{|U|,max}\}$$

Let $E(n_i,u_k, f_{k,h})$ represent the dynamic consumption ofenergy in task $n_i$on the $VM_{uk}$with frequency $f_{k,h}$. This is calculated by

$$E(n_i, u_k, f_{k,h}) = (P_{k,ind}+ C_{k,ef}\times f_{k,h}^{mk}) \times W_{i,k} \times f_{k,m}$$

Considering the energy utilization of the DAG application, Gisthe total energy consumption of all jobs, E(G) calculated by

$$E\ (G) =\sum_{i=1}^{|N|} E(n_i)=\sum_{i=1}^{|N|} E= (n_i,u_{pr(i)},f_{pr(i)},h_{z(i)}),\ldots\ldots.(4$$

$$Cost\ (G) =\sum_{i=1}^{|N|} cost(n_i)=\sum_{i=1}^{|N|} cost(n_i, u_{pr(i)}, f_{pr(i)}, h_{z(i)})\ldots$$

Where$u_{pr(i)}$and $f_{pr(i)}$,$h_{z(i)}$represent the assigned VM and frequency of n_i, respectively.

5.3 Cost Budget Scope

High frequency implieslow arrival time. Lowest frequency implies the high arrival time for a task on the similar VM.We can obtain the highand low costs of task n_i, represented by $cost_{min}(n_i)$ and $cost_{max}(n_i)$, in that order, by traversing all the VMs

$$cost_{min}(n_i) = \min u_k \text{ € } U\ cost(n_i, u_k, f_{k,max}) \quad \ldots\ldots.(6)$$

and

$$cost_{max}(n_i) = \max u_k \text{ € } U\ cost(n_i, u_k, f_{k,low}) \quad \ldots\ldots.(7)$$

Then, the total of each task, and the high and low costs of the DAG application respectively are the total cost of the DAG application G.

$$cost_{min}(G) =\sum_{i=1}^{|N|} cost_{min}\ (n_i) \quad \ldots\ldots.(8)$$

and

$$cost_{max}(G) =\sum_{i=1}^{|N|} cost_{max}\ (n_i) \quad \ldots\ldots.(9)$$

The cost budget of the DAG program$cost_{budget}(G)$ must be bigger than or equal to $cost_{min}(n_i)$; otherwise, $cost_{budget}(G)$ is always unsatisfied. Meanwhile, $cost_{budget}(G)$ must be lesser than or equivalent to $cost_{max}(n_i)$; otherwise, $cost_{budget}(n)$ is everytimecontented. Therefore, we can come to an assumption that $cost_{budget}(G)$ belongs to the scope of $cost_{min}(G)$ and $cost_{max}(G)$:$cost_{min}(G)) \leq cost_{budget}(G) \leq cost_{max}(G)$.

TEST Engineering & Management

January - February 2020
ISSN: 0193 - 4120 Page No. 6056 - 6070

VI. AIM:

The aim of the paper is to select the optimal Virtual Machine for Energy efficient workflow scheduling using RECABP (Reduced Energy Consumption using the Available Budget Pre-assignment Algorithm) through techniques of Pre-assignment of cost and energy consumption.

6.1 An Overview of Proposed Reduced Energy Consumption using the Available Budget Pre-assignment (RECABP) Algorithm:

Reduced Energy Consumption using the Available Budget Pre-assignment (RECABP) Algorithm is proposed. In term of the proposed system, we suggest the consumption of energy with cost budget mechanism.As a result, thePhysical Machines (PM) in a heterogeneous system can be able to differentiate between energy scheduling and energy efficiency in lower bound users in such a way that they do not share resources. A robust RECABP in the heterogeneous cloud will focus on multitenancy. Multi-Tenancy introduce to assets resource sharing in Cloud Computing where any resource device is reusable in the Cloud infrastructure. In a multi-tenant cloud, a user depends on the PM for trustworthy co-tenants VM. We include a novel reputation management mechanism that encourages the VM to assign good co-tenants to a good user.

In this paper, we highlight the energy methodical scheduling using Reduced Energy Consumption using Available Budget Pre-assignment Algorithm(RECABP) for cloud scheduler to is basically minimize the energy consumption.The main aim of RECABP is that the cost budget of the given program is transmitted for every workflow by pre-assigning available budget Pre-assignment counts to VM'S which are not assigned workflows. An available budget has been proposed by using pre-assignment method to improve the existing systems.

6.2 Available Budget Pre-assignment

The use of the available budget pre-assignment methodis to reduce the energy utilization and Cost budget.

To satisfy Cost budget in this stage, the cost budget of the application is transferred to each task by availing the existing budget pre-assignment.

In the next step, in order to reduce energy consumption,we allow each task to select thecombination enterprise of VM and frequency that has the least energy utilization while satisfying the cost budget of the task.

Available energy–AB(G) is the available budget between the overall cost of the application as allotted by the budget-$cost_{budget}(G)$ and minimum cost -$cost_{min}(G)$ of the given DAG application G. Available energy–AB(G) is computed as the difference between the Cost of the overall budget and the minimum Cost of the application.

$$AB(G) = cost_{budget}(G) - cost_{min}(G) \quad …….(10)$$

On Comparing with the budget level pre-assignment, the vital development in this proposed systemis that the pre-assignment of $n_{s(z)}$ is changed from $cost_{bl}(n_{s(z)})$ to $cost_{avail}(n_{s(z)})$, and thus, $cost_{bl}(n_{s(z)})$ is changed to$cost_{avail}(n_{s(z)})$ in computing $cost_{budget}(n_{s(y)})$

$$cost_{avail}(n_{s(z)}) = (G) - \sum_{x=1}^{y-1} = cost(n_{s(x)}) -$$
z=y+1|N|costavail(ns(z))…….(11)

$cost_{avail}\ (n_{s(z)})$is computed by

$$cost_{avail}\ (n_{s(z)}) = \frac{cost_{min}\ (n_{s(z)})}{cost_{min}\ (G)} \times AB(G) + cost_{min}\ (n_{s(z)}) \quad …….(12)$$

Where $cost(n_{s(x)})$ is the actual used cost of $n_{s(x)}$ and $cost_{min}(n_{s(z)})$ is the pre-assigned cost of $n_{s(z)}$. The

January - February 2020
ISSN: 0193 - 4120 Page No. 6056 - 6070

budget level pre-assignment is that the budget pre-assignment value of $n_{s(z)}$ is shifted to $cost_{bl}(n_{s(z)})$.

RECABP Algorithm:
Input: G= (N, W, M, C), U, $cost_{budget}(G)$
Output: E(G), cost(G) and its related values
Step1: Each task is prioritized in a Virtual Machine using rank_list
Step 2: While (rank_listis not null) do
Step 3: rank_list. out() values are assigned to $n_{s(y)}$;
Step 4: $n_{s(y)}$ value is assigned n;
Step 5: Compute$cost_{min}(G)$by using Eq. (7);
Step 6: Compute$cost_{budget}(t_i)$by using Eq. (10);
Step 7: $E(n_i) = +\infty$;
Step 8: For (each VM_{uk}€ U) do
Step 9: For (each frequency $f_{k,h}$in [$f_{k,low}$ and $f_{k,max}$]) do
Step 10: Computecost (n_i, u_k,$f_{k,h}$) for the task t_i;
Step 11: if (cost(n_i, u_k, $f_{k,h}$) $\leq cost_{budget}$(ti)) then
Step 12: ComputeE(n_i, u_k, $f_{k,h}$) ;
Step 13: if (E(n_i, u_k, $f_{k,h}$) <E(t_i)) then
Step 14: E(n_i) $\leftarrow$(n_i, u_k, $f_{k,h}$);
Step 15: cost(n_i),$\leftarrow$cost(n_i, u_k, $f_{k,h}$);
Step 16: end if
Step 17: break;
Step 18: end if
Step 19: end For
Step 20:end For
Step 21: end while
Step 22: Compute Available Budget AB (G) by using Eq. (10);
Step 23: Computecost(G)by using Eq. (5);
Step 24: ComputeF(G)by using Eq. (4);

The central idea of RECABP is that the cost budget of the DAG application is shifted to each task by preassigning available budget preassignment values to unassigned tasks.

a. Each task is prioritized by RECABP based on $rank_u$ values in Step 1.
b. RECABP allocates every task to the VM iteratively with the reduce energy consumption of the DAG application in Steps 2-21.
c. RECABP calculates the cost budget of n_i by using equation in Steps 4-6.
d. While satisfying its cost budget, the effective combination assignmentof VM and frequency is selected by RECABP for $n_i$in Steps 8-20.
e. In Steps 22 and 23, the related values of the DAG application are calculated.

$O(|N|^2 * |U| * |E|)$ is the time complexity of the RECABP algorithm. The details are evaluated as follows:

(i) The sum of all those jobs, which can be performed within O(|N|) time is calculated as the energy utilization of the application. (the While loop in Steps 2-21).

(ii) O(|N|* |U|*|E|) is the time taken for selecting the combination assignment of VM and frequency (the For loop inSteps 8-20). |E| refers to the maximum frequencies number on the VM.

Which areu_{pr}(i) and f_{pr}(i), hz(i) represents the required VM and frequency of $n_i$respectively.
Explanation of the Algorithm:
This algorithm reduces the energy consumption. As a result, itgains an apt frequencyassignment. Further, itsolves the problem of reducing expenditure. The main advantage is that the algorithm pre-assigns the available Budget. The $rank_u$is computed based upon the nodes assigned to the task. The more number of nodes assigned to the task, the $rank_u$ will be higher. Accordingly, RECABP assigns the cost budget based upon the tasks assigned. In addition, it selects the effective combination of virtual Machines according to the cost Budget. Finally,it calculates the value of the DAG application.

January - February 2020
ISSN: 0193 - 4120 Page No. 6056 - 6070

VII. ADVANTAGES OF RECABP ALGORITHM

- The issue of minimizing energy consumption in heterogeneous cloud computing systems for cost budgeted application is addressed.
- The schedule length is reduced without involvingconsumption of energy.
- The consumption of energyof the DAG programis minimised and the cost budget is also satisfied.

VIII. METRICESCONSIDERED

8.1 Cost Function

The task specifications are defined as either in invariant form or dynamic. The invariant form is fixed for the whole lifetime of the task. In the dynamic specification, random variation at time of execution.

To find out optimal placement of jobs, the cost function of Available Budget AB(G) is considered. Accessible energy is computed as the difference available budget between $cost_{budget}(G)$ and $cost_{min}(G)$ for a given node 'G'.

$$AB(G) = cost_{budget}(G) - cost_{min}(G)$$

8.2Time Preference (TP):

Time Preference shows the need of selecting instances of stock assets. Time Preference gives Workflows the least execution time.
TP_igiven as

$$TP_i = \frac{TR_i}{TR_{max} - TR_{min} + 1}$$

Where $TR_i = T_iF\ in - T_iStart$, T_iF in is the finish time of the i^{th}Workflow, $T_iStart$is the start time of the i^{th}Workflow.$TR_{max}$is time taken by slots Workflow, $TR_{min}$is time taken by fastest Workflow. $TR_i$is execution time spent in the flow of work in a queue manner.

8.3 Bandwidth Preference (BP):

It indicates the sole purpose of selecting instances of resources that provide the appropriate bandwidth for Workflows. $BP_i$is computed as

$$BP_i = \frac{BR_i}{BR_{maxx} - BR_{min} + 1} \quad ; 0 \leq bwtypeLIST.size()$$

$$BR_i = BW_{vm.getbw(i)} bwtypeList.size()$$

This equation is the count of Workflows reduces bandwidth, $BR_{max}$is maximum bandwidth in flow of work, $BR_{min}$is the minimum demanded bandwidth in the flow of work, $BR_i$is bandwidth taken by executing i^{th}Workflow in the Workflow queue.

8.4Energy Consumption:

The energy consumed is calculated using formula

$$E(G) = \sum_{i=1}^{|N|} E(n_i) = \sum_{i=1}^{|N|} E = (n_i, u_{pr(i)}, f_{pr(i)}, h_{z(i)}),$$

where $u_{pr}(i)$,$f_{pr}(i)$ andhz(i) represent the assigned VM and$n_i$respective frequency.

IX. 9.**COMPLEXITIES** INVOLVED IN PROPOSED SYSTEM

- Uncertainty: By predicting task execution time and waiting time to improve resource utilization and efficiency.
- Quality of Service: Service Provider has to ensure that services must be delivered with minimum Qos.
- Load Balancing: The optimal utilization of cloud resources demands uniform load must be distributed among different Virtual Machines.

X. EXPERIMENTS AND EVALUATION

10.1Experimental Environment

For the performance evaluation of our RECABP algorithm, thewidely-used CloudSimframework is adopted as the cloudsimulation environment for the

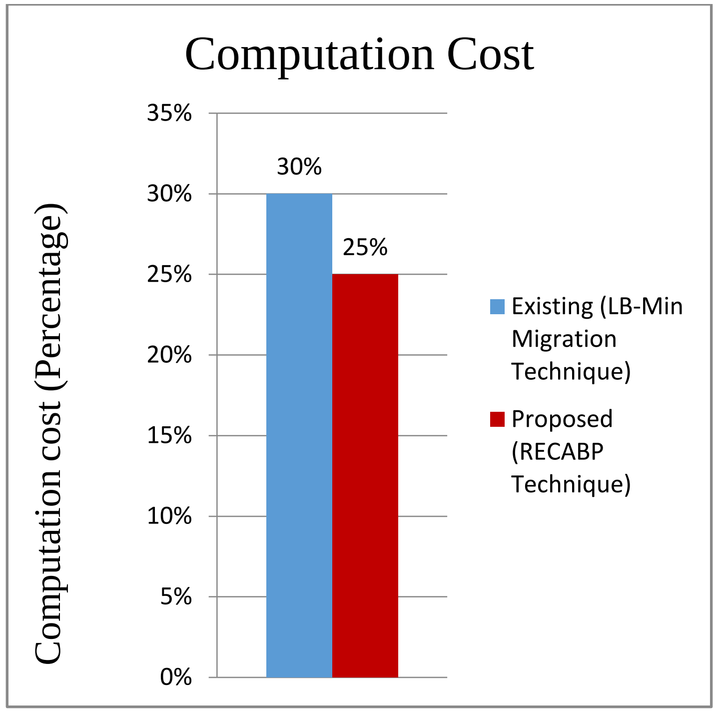
January - February 2020
ISSN: 0193 - 4120 Page No. 6056 - 6070

experiments. Theexperimental setup aims at minimization of energy consumption during workflow in the cloud computing. The proposed a Reduced Energy Consumption using the Available Preassignment (RECABP) algorithm is to reduce energy consumption. First, the local resource fetches the resources needed by the end user. The resources are then allocated by assigning the available resources in an efficient way. The optimal virtual machines are created and the cost is estimated. Task manager determines the scheduling sequence and resource assignments for the requests and allocates suitable resources to each workflow under the help of the scheduling algorithm.

10.2Graph Design Based Result

An essential way of displaying a result is plotting a graph. So, we use agraphed sign-based result to show a result comparison with light weight feature classification data selection between Lower Bound Min Migration and RECABP Algorithm. We will prove that Energy Consumed by proposed algorithm is reduced.The performance comparison of RECABP Algorithm and Lower Bound-Min-Migration under system of differentconstraints during scheduling workflowson cloud computing is highlighted in the graph.

The performances of the existing and proposed technique are depicted in Graph-1. It indicates the comparison of Computation Cost between the existing and the implementedproposedtechnique. Graph-2 represents the comparison of Execution time between theexisting and the proposed technique. Graph-3 highlights the comparison of Band-with Preference betweenthe existing and the proposed technique implemented. Graph-4 depicts the difference in Energy Consumption between the existing and proposed technique implemented.



Graph 1: Comparison of Computation Cost in both Existing and Proposed techniques

Significance of Graph 1:

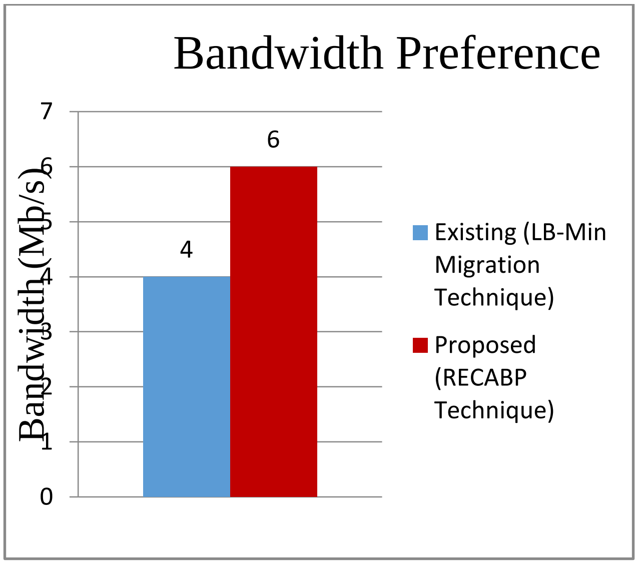
The Graph 1 represents the reduced percentage of Computation cost. In the above Graph, the Algorithm is represented by X-axis and Y-axis shows the Computation cost in terms of percentage. The proposed work is approximately 5-10% greater than the existing technique.

Graph 2: Comparison of Execution time in both Existing and Proposed techniques

Significance of Graph 2:

The Graph 2 shows the percentage of time that is executed. In the above Graph, X-axis depicts the Algorithm and the Execution Time in seconds is denotes in Y-axis. The proposed technique is approximately 0.2 seconds better than the existing technique for all the three datasets. This fact is highlighted in the above Graph. Better performance of the proposed technique with respect to the Existing technique is clearly depicted in the Graph 2.

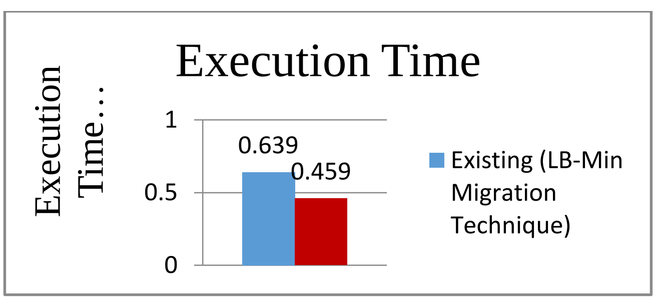
January - February 2020
ISSN: 0193 - 4120 Page No. 6056 - 6070

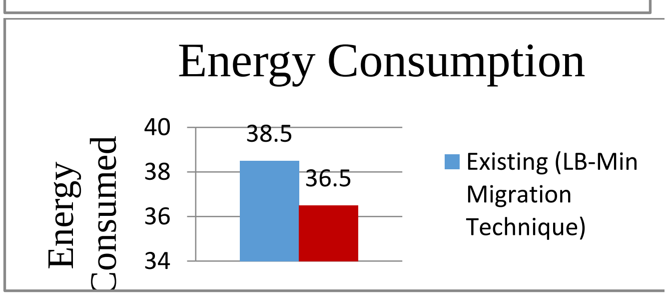


Graph 3: Comparison of Bandwidth Preference in both Existing and Proposed techniques

Significance of Graph 3:

The effective increase in the percentage of Bandwidth is depicted in Graph 3. In the above Graph, X-axis denotes the Algorithm and Y-axis depicts the Bandwidth. The result of the proposed technique is approximately 10-15% better than the existing technique.





Graph 4: Comparison of Energy Consumption in both Existing and Proposed techniques

Significance of Graph 4:

The Graph 4 shows the decrease in the percentage of Energy consumed during workflow scheduling for the given dataset. In the above Graph, X-axis denotes the Algorithm and theEnergy Consumed is represented by Y-axis. The proposed technique is approximately 2-5% better than the existing technique forthe given dataset. This fact is highlighted in the above Graph. The better performance of the proposed work with respect to the existing work is clearly depicted in the Graph 4.

XI. CONCLUSION

The problem ofminimizing energy consumptionin scheduling workflows is solved by using RECABP Algorithm. RECABP attempts to reduce consumption of energy. Further, the cost budget of the Big Data devices is satisfied by the method of available Budget Pre-assignment. RECABP Algorithm overcomes the complexity involved in Uncertainty, Scalability, and Quality of Service in workflow scheduling. RECABP comparatively elucidates DAG applications with its effects in the existing algorithms through experiments.

We will incorporate more energy efficient aspects into our algorithms in the future to improve their further frequency. In the future, the method of dynamic resource allocation can be used to handle heterogeneous workloads and validates by working on the Big Data environment.

XII. REFERENCES

[1] Ping Kuang, WenxiaGuo, Hongjian Li, WenhongTian, RajkumarBuyya, "Analyzing Energy-Efficiency of Two Scheduling Policies in Compute-intensive Applications on Cloud", IEEE Translations and Data Mining, 2018.

[2] Yiping Wen, Jianxun Liu, Wanchun Dou, XiaolongXu, Buqing Cao and Jinjun Chen, "Scheduling workflows with privacy protection constraints for Big Data applications on cloud", Future Generation Computer Systems 2018.

[3] Shridhar G. Domanal, G. Ram Mohana Reddy, “An efficient cost optimized scheduling for spot instances in heterogeneous cloud environment”, Future Generation Computer Systems, 2018.
[4] Tingming Wu, HaifengGu, Junlong Zhou, Tongquan Wei, Xiao Liu, Mingsong Chen, “Soft Error-Aware Energy-Efficient Task Scheduling for Workflow Applications in DVFS-Enabled Cloud”, Journal of Systems Architecture, 2018.
[5] EhabNabielAlkhanak, Sai Peck Lee, “A hyper-heuristic cost optimisation approach for Scientific Workflow Scheduling in cloud computing”, Future Generation Computer Systems, 2018.
[6] Vishakha Singh, Indrajeet Gupta, Prasanta K. Jana, “A novel cost-efficient approach for deadline-constrained workflow scheduling by dynamic provisioning of resources”, Future Generation Computer Systems, 2017.
[7] N. Moganarangan, R.G. Babukarthik, S. Bhuvaneswari, M.S.SaleemBasha, P.Dhavachelvan, “A novel algorithm for reduced energy consumption in cloud computing environment: Web service computing approach”,Journal of King Saud University – Computer and Information Sciences 28, pp. 55–67, 2016.
[8] Juarez, F., Ejarque, J., Badia, “Dynamic energy-aware scheduling for parallel task-based application in cloud computing”, Future Generation Computer Systems 78, pp: 257– 271, 2016.
[9] Zhaomeng Zhu, Gongxuan Zhang, Miqing Li, Xiaohui Liu, “Evolutionary Multi-Objective Workflow Scheduling in Cloud”, Transactions on Parallel and Distributed Systems, 2016.
[10] Zhongjin Li, JidongGe, Haiyang Hu, Wei Song, Hao Hu, and Bin Luo, “Cost and Energy Aware Scheduling Algorithm for Scientific Workflows with Deadline Constraint in Clouds”, IEEE Transactions on Services Computing, 2015.
[11] 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, 2015.
[12] R.N. Calheiros, R. Ranjan, A. Beloglazov, C.A.F. De Rose, R. Buyya, CloudSim:atookit for modeling and simulation of cloud computing environments and evaluation of resource algorithms, Softw. - Pract. Exp. 41 (1) (2011) 23–50.