

Intelligent Workload Management in Virtualized Cloud Environment

R.Sujitha¹

¹SriGuru Institute of Technology, CSE,
Sujitharaju4@gmail.com

N.VijayaRaghavan²

²SriGuru Institute of Technology, CSE
nvijayms@gmail.com

Abstract— Cloud computing is a rising high performance computing environment with a huge scale, heterogeneous collection of self-sufficient systems and elastic computational design. To develop the overall performance of cloud computing, through the deadline constraint, a task scheduling replica is traditional for falling the system power utilization of cloud computing and recovering the yield of service providers. To improve the overall act of cloud environment, with the deadline constraint, a task scheduling model is conventional for reducing the system performance time of cloud computing and improving the profit of service providers. In favor of scheduling replica, a solving technique based on multi-objective genetic algorithm (MO-GA) is considered and the study is determined on programming rules, intersect operators, mixture operators and the scheme of arrangement of Pareto solutions. The model is designed based on open source cloud computing simulation platform CloudSim, to obtainable scheduling algorithms, the result shows that the proposed algorithm can obtain an enhanced solution, thus balancing the load for the concert of multiple objects.

Index Terms— Cloud Computing, CloudSim, Deadline, Multi-Objective Genetic Algorithm, Task Scheduling.

◆

1 INTRODUCTION

Cloud computing is a recently successful area and has been rising as a marketable veracity in the information technology field. It is a computing paradigm, which provides computing as a service based on internet application. Cloud computing provides infrastructure, platform, and software (application) as services, which are made presented as contribution based services in a pay-as-you-go model to clients and these computing services are delivered to the users through the Virtualization Technology. In cloud application, delivery time and cost are important aspects, so the delivery service will be provided based on a certain time limit which creates the deadline to the provider, where deadline depends on task completion. Deadline allows user specify a job's deadline and tries to formulate the job be finished earlier than the deadline. During the job deadline, we can build a model to proceed the reality of the task enduring time estimating in the heterogeneous situation, put together the use jobs can be finished before the deadline extreme. Reasonably, the foremost demand of cloud computing is to facilitate the customers only utilize what they require, and only pay for what they really apply. Resources are presented to be accessed, since the cloud at any particular time, and from any location through the internet [7]. Yet, data centers use a considerable and rising portion of energy; a regular data center consumes as much energy as 25,000 households. Hence, energy-aware computing is critical for cloud computing systems that consume significant quantity of energy.

2 RELATED WORK

The resource stress for diverse jobs alters over time. Job scheduling system, which capably allocates resources to necessary tasks under the restriction of the Service Level Agreements (SLAs), is a fundamental concern in achieving soaring act in cloud computing and of large consequence for getting better resource load balance, defense, consistency and sinking energy utilization of the Entire system. However, it is a huge demanding problem for competent cloud computing setting. Towards reduce the energy consumption, Pinheiro et al. Propose a model for minimization of power consumption in a various cluster of computing nodes allocation several web-applications, which repeatedly monitors the load of resources and makes decisions on switching nodes on/off to play down the generally power consumption [8]; Raghavendra et al. mix five diverse power supervision policies and discover the problem in conditions of manage theory, but the system fails to maintain variable SLAs for dissimilar applications [9]; Lee et al. propose two algorithms depends on pricing replica, via processor contribution in order to balance among profit and resource consumption [10]; Garg et al. propose a linear programming focused genetic algorithm, aiming to ascertain the most excellent scheduler in a utility grid by minimizing the collective costs of every single one users in a corresponding method [11].

Each and every one of the above mentioned methods believe the profit or the energy in their study, except do not the affiliation among them. To conquer the deficiencies of the beyond algorithms, in this paper, we first ascertain a macroscopic scheduling replica through cognition and assessment workings for the cloud computing, which considers together the desires of different jobs and the situation of computing communications, then propose a job scheduling algorithm based on Multi-Objective Genetic Algorithm (MO-GA), captivating into account of the energy consumption and the profits of the service providers, and given that a dynamic mixture system of the majority suitable scheduling scheme for users according to the real-time desires; at preceding, we take several experiments to certify our design and measure up to our MO-GA based scheduling replica to the usual ones.

3 MODEL FOR JOB SCHEDULING

In cloud computing, service requirements have heterogeneous resource anxiety as some services might be CPU demanding while others are I/O-intensive. Cloud resources want to be allocated not only to convince Quality of Service (QoS) requests to specific by users through SLAs, but moreover to diminish energy convention and get better the profits to the service Providers. The scheduling replica we include recognized the feature functions of the main apparatus are introduced as follows: apply for cognition factor should be fully sensitive of the particular requirements for unlike businesses, which may perhaps take in the computing, storage and communication wishes for computing, advent law and synchronized conditions, security and privacy desires, QoS of the service and so scheduled; Service decay module decomposes the service demand into different stage of granularities among different processor preferences. In the subsequently procedure, the task administrator will analyze the resource requests of every granularity, and mapping it on top of an optimal processors to arrive at an effectual solution. Task manager is conscientious for task position management (start, stop, cancel...), formative the scheduling series and resource handing over for the requests and allocating apposite resources to apiece job under the assist of the scheduling algorithm. Resource cognition module plays the position of supervision the existing resources, monitoring the performances of assets, dynamic optimization of scheduling scheme and error announcement.

4 PROBLEM FORMULATION

In our representation, a cloud application is considered as a set of work items or jobs to carry out a multifaceted computing task via using cloud resources, and the set $A = (a_1, a_2, \dots, a_m)$ is a consignment of applications arrived during a period. Throughout the scheduling process, the client yields a service request in favors of application $a_i (1 \leq i \leq M)$, through the resource desires characterized as a leash (t_i, n_i, d_i) where, t_i represents the hesitation time of the application for virtual machines (VMs), which are the virtualized calculating elements in cloud computing by means of virtualization technology, n_i for the number of VMs essential for a_i and d_i for the deadline following what the application will be measured to be abortive.

The difficulties require to solve for this algorithm is how to schedule these M applications to the certain N clouds under the constraints and compose the objective function most select. Where, the N clouds distributed in dissimilar geographical areas around the world are typically heterogeneous, whereas in a cloud, all the VMs are well thought-out homogeneous with the virtualization techniques.

4.1 Objective Function

Assume application a_i is scheduled to accomplish on cloud c_i , and p_i characterized the Power of every VM in c_i , then, the energy consumption for execution of a_i is given by

$$E_{ij} = p_i n_i t_i \tag{1}$$

Along with the profit of the service contributor is:

$$R_{ij} = n_i t_i p_r - c_{oij} \tag{2}$$

Where, the p_r is the price unit stimulating by supplier for application a_i , and c_{oij} is the cost of the provider for completing the application a_i . Where, E and R is the whole energy utilization and profit for the execution of M application on N clouds correspondingly.

4.2 Limitations

The limitations are planned as follows:

- (1) The application a_i has to be completed before the deadline d_i , or else, the schedule is measured to be abortive;
- (2) Each one application can be present and allocated to only one cloud.

5 MO-GA SCHEDULING ALGORITHM

5.1 Encoding Rule

Each one schedule and articulated as a 2 by M matrix, where, M is the length of the chromosome. The first row of the matrix symbolized the demand applications, and second of the matrix is the equivalent number of the cloud where the application is performed.

Fig. 1 shows an example of scheduling result, in which, application 2 is allocated to cloud 0, and application 1 is allocated to cloud 5.

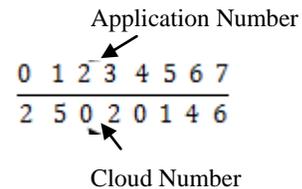


Fig.1. Encoding example of a Scheduling

Pursuant to the above rule, we can see with the intention of each application can only be consigned to one cloud, while a cloud may perhaps capable to process numerous applications.

IJTET©2015

(1) Individual Evaluation

In this paper, the fitness is deduced from the energy consumption and profits of the service providers. Only the solutions with the most excellent rank after the evaluation of the fitness function are stored in the Pareto documentation which contains the altered non-dominated solutions generated during the generations.

(2) Selection operation

The selection operation is based on contest operator of k individuals, with two strategies: superiority and crowding. The superiority policy makes use of the individuals in Pareto archive and selects the best ones according to the non-dominated model to the subsequently generations, allowing the junction of the evolution method. Crowding strategy takes advantage of crowding distance to estimate the strength of surrounding solutions and take out the solutions which were too crowded by ranking the crowding distance of each one individual. The crowding expanse is defined as the fringe of the rectangle defined by its left and right neighbors, and infinity if there is no neighbor.

(3) Crossover Operation

The crossover operator brings into play two individuals s_1, s_2 to makes two new individuals s'_1, s'_2 . For individual s_1 , first, the operator arbitrarily produces two integers i, j , where, $1 \leq i \leq j \leq N$; then, replicas the tasks in s_1 before i and after j to s'_1 , and maps the tasks between i and j to a transitory individual s_1^m according to the tasks distribution result in s_2 ; finally, copies the tasks in s_1^m to consequent place in s'_1 , as shown in Fig. 3. The individual s'_2 is generated using the same method.

(4) Mutation Operation

The mutation operation desires two tasks in a being randomly, and switches their allocation position to makes a recent individual.

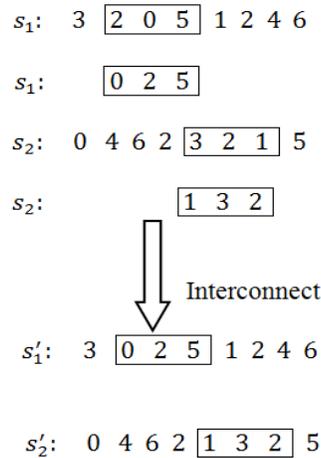


Fig. 3. Crossover operation mechanism

5.2 Population Initialization

The population initialization involved the quality of the prospected generations, and is an significant step in the whole algorithm. In this paper, this step is accomplished by combing the arbitrary and greedy initialization methods. Owing to the greedy beginning method, the scheduler discards the applications not assembly the deadline constraint which may cause the entire scheduling disastrous. This type of initialization scheme helps add variety to the initial population and let alone biasing the seeking of MO-GA.

5.3 Genetic Algorithm

Genetic algorithm is a search heuristic that perform the process of natural development based on a population of candidate solutions. It is usually used to create useful solutions to optimization and problems. Produce an initial population by randomly generated individual. In the process of progress, a modification is performed by those operators on each creature.

- Evaluate the fitness of all individuals
- While termination condition not met do
 - Select fitter individuals for reproduction
 - Crossover between individuals
 - Mutate individuals
 - Evaluate the fitness of the modified individuals
 - Generate a new population
- End while

Each chromosome represents a scheduling result, and an evaluation operator (fitness) is called to evaluate the issue.

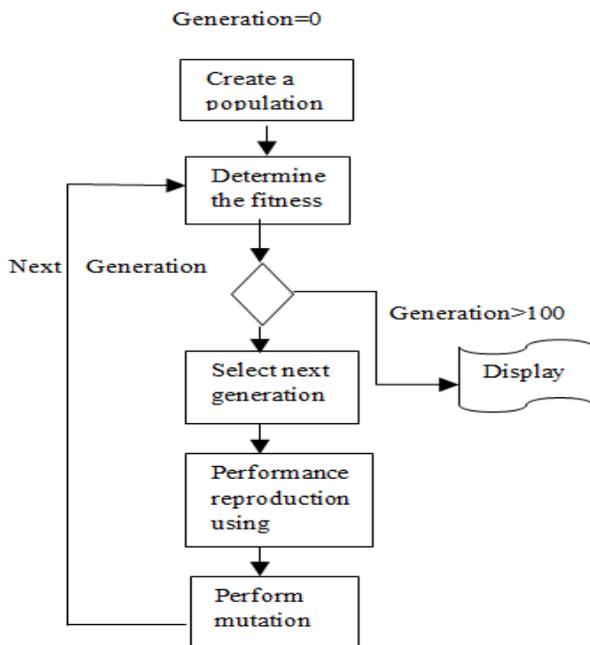


Fig. 2. Flowchart for genetic algorithm

5.4 Optimal Selection in Pareto Archive

The grades of MO-GA algorithm are locates of Pareto solutions, on condition that an extensive range of possible alternatives, whereas tumbling the efficiency of scheduling process. In exercise, users now and then need to adjust the degree of favorite for a particular intention dynamically. This measure provides an advance to pick up an optimal solution along with the external Pareto archive according to the recent requirement. A two dimensional vector is establish to correspond to the weighting for a particular objective, whose direction points to the most approving solution.

Fig. 4 shows an example with 3 two-dimensional vectors, where, $P_1 - P_5$ signify the external collection of after the MO-GA algorithm, $v_1 = (0,1)$, $v_2 = (\sqrt{2}/2, \sqrt{2}/2)$ and $v_3 = (1,0)$ represents three kind of desires respectively.

For example, P_1 is the optimal solution used for vector v_1 , and P_3 for v_2 , P_5 for v_3 .

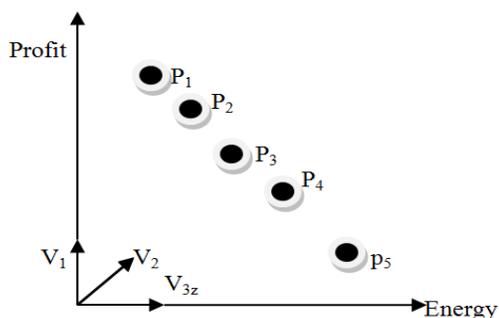


Fig. 4. The schematic diagram of optimal selection

6 IMPLEMENTATION

Genetic Algorithm obtains terminated after user specified number of generations. To find the better results, it's evaluated 30 genetic results. Based on the above results, the implementation steps of this algorithm are listed following:

- (1) Early the population by greedy and random technique;
- (2) Transform the individual for the duration of the evolution process of the MO-GA algorithm according to the operators designated and store the results to peripheral Pareto archive;
- (3) Go for the optimal solution according to the vector and realize the scheduling result to distributed cloud confederacy.

7 CONCLUSION AND FUTURE WORK

Clouds enable the users to use utility services. Users are required to pay for access to the services based on their usage and level of quality of service required. In this research we have proposed a modified genetic algorithm for single user jobs in which the fitness is developed to encourage the arrangement of solutions to achieve the time minimization. In this paper, establishing a scheduling model for cloud computing based on MO-GA algorithm to minimize energy consumption and maximize the profit of service provides under the constraint of deadlines.

We first propose a job scheduling architecture under the environment of cloud computing, which contains several components to analyze the application, and allocate the suitable resources to the applications to improve the effectiveness and efficiency of the computing; then, the MO-GA based scheduling algorithm is proposed, at last, several experiments are conducted to validate our scheduling models. In future, we have to enhance the algorithm by supporting runtime scheduling which is considering the user's quality of service and priority of jobs for multiple users.

REFERENCES

- [1] Armbrust M, Fox A, Griffith R, Joseph A D, Katz R, Konwinski A, Lee G, Patterson D, Rabkin A and Stoica I, "A view of cloud computing", Communications of the ACM, Vol. 53, No. 4, 2010.
- [2] Nidhi Jain Kansal and Indrveer Chana, "Cloud. Load Balancing Techniques: A Step Towards Green Computing", IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 1, No. 1, 2012.
- [3] Iosup, A., Ostermann, S., Yigitbasi, M.N., Prodan, R., Fahringer, T. and Epema, D.H.J, "Performance Analysis of Cloud Computing Services for Many-Tasks Scientific Computing", IEEE Transactions on Parallel and Distributed Systems, Vol. 22, No. 6, 2011.
- [4] Arunadevi.M and R.S.D Wahidabanub, "Design of Power Efficient Schema for Energy Optimization in Data Center with Massive Task Execution Using DVFS", IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 1, No 2, 2012.
- [5] Almutairi, A., Sarfraz M., Basalamah S., Aref W. and Ghafoor A, "A Distributed Access Control Architecture for Cloud Computing", IEEE Software Vol. 29, No. 2, 2012.
- [6] Junaid Qayyum, Faheem Khan, Muhammad LaL, Fayyaz ul, Muhammad Sohaib and Fahad Masood, "Implementing and Managing framework for PaaS in Cloud Computing", IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 5, No. 3, 2011.
- [7] Sanjeev Narayan Bal, "Clouds for Different Services", IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 4, No 1, 2012.
- [8] E. Pinheiro, R. Bianchini, E.V. Carrera and T. Heath, "Load balancing and unbalancing for power and performance in cluster-based systems", in Proceedings of the Workshop on Compilers and Operating Systems for Low Power, 2001.
- [9] R. Raghavendra, P. Ranganathan, V. Talwar, Z. Wang and X. Zhu, "No "power" struggles: coordinated multi-level power management for the data center", SIGARCH ComputerArchitecture News, Vol. 36, No. 1, 2008.
- [10] Lee Y.C., Wang, C., Zomaya, A.Y. and Zhou B.B., "Profitdriven service request scheduling in clouds", In: Cluster, Cloud and Grid Computing (CCGRID), 2010.
- [11] Garg, S.K., Konugurthi P. and Buyya R, "A linear programming driven genetic algorithm for meta-scheduling on utility grids", International Journal of Parallel Emergent and Distributed Systems, Vol. 26, No. 6, 2011.bonafide