Optimizing Energy Consumption in Clouds by using Genetic Algorithm

Sahar hosseinzadeh[1]
Department of Computer Engineering, Sari Branch, Islamic Azad University
Sari, Iran

s.hoseinzadeh93@gmail.com

Mirsaeid hosseini shirvani[2*]
Department of Computer Engineering, Sari Branch, Islamic Azad University
Sari, Iran

Corresponding Author e-mail:

mirsaeid_hosseini@yahoo.com

Abstract— cloud computing has recently become popular technology. Much of the literature shows that energy consumption and resource utilization in clouds are highly coupled. To integrate and make good use of resources at various scales, cloud computing needs efficient methods to manage them .In this paper, a genetic algorithm was proposed to efficiently allocate tasks to virtual machines, which allocates resources based on available resources and the energy consumption of each virtual machine. Evaluation results show that the proposed algorithm has more scale up and less energy consumption than first-fit decreasing (FFD) and best-fit decreasing (BFD) algorithms.

Keywords—Virtual machine; Data center; Cloud computing; Genetic algorithm; Energy consumption

I. INTRODUCTION

With the rapid development of social information level, the dependence on information systems in various industries, including government, economy and so on, is more and more seriously. Since IBM announced the cloud computing program in the end of 2007, cloud computing concept appeared in public. The concept of cloud computing that people generally accepted is defined by National Institute of Standards and Technology (NIST)[11]: Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. To fully realize the potential of the Cloud computing, Cloud service providers have to ensure that they can be flexible in their service delivery to meet various consumer requirements. Until recently, high performance has been the sole concern in data center deployments, and this demand has been fulfilled without paying much attention to energy consumption. Although, increasing energy costs cause dwindling availability. Also, higher power consumption results not only in boosted electricity bills, but also in additional requirements to a cooling system and power delivery infrastructure. Another rising concern is the environmental impact in terms of carbon dioxide (CO2) emissions caused by high energy consumption. Therefore, in the design of modern computing systems, there is a need to shift the focus from optimizing data center resource management for pure performance to optimizing them for energy efficiency [1], [2].

II. RELATED WORK

Garg et al. in [4] have investigated the problem of energy and CO2 efficient scheduling of HPC applications in geographically distributed Cloud data centers. The authors have proposed five scheduling policies, two of which minimize carbon dioxide emissions, two maximize the profit of resource providers, and the last one is a multi-objective policy that minimizes CO2 emissions and maximizes the profit. The energy consumption is also reduced by applying dynamic voltage frequency scaling (DVFS) technique for all the CPUs in data centers [12].The experimental results have shown that the energy-centric policies allow the reduction of energy costs by 33% on average. Much of the literature shows that energy consumption and resource utilization in clouds are highly coupled. On the other hand, due to the inherent characteristics of cloud computing, the resources are controlled by the cloud providers and thus can be managed more efficiently.

Buyya et al. in [5] have proposed the GreenCloud project aimed at development of energy-efficient provisioning of Cloud resources, while meeting quality of service (QoS) requirements defined in SLA established through a negotiation between providers and consumers. The authors have proposed three policies for scheduling real-time virtual machines (VMs) in a data center using DVFS to reduce the energy consumption, while meeting deadline constraints and maximizing the acceptance rate of provisioning requests. The proposed approach has been evaluated via simulations using the CloudSim toolkit.

Kim et al. in [6] suggested a model without dedicated measurement hardware to estimate the energy consumption of a virtual machine based on in-processor events generated by the virtual machine. Based on this estimation model, a virtual machine scheduling algorithm is proposed to provide computing resources according to the energy plan of each virtual machine. The suggested scheme was implemented in the Xen virtualization system, and the evaluation shows that the suggested scheme estimates energy consumption and accordingly provides computing resources with errors of less than 5% of the total energy consumption. The suggested scheme can be used as a billing basis for cloud systems.

Stillwell et al. in [7] have studied the problem of resource allocation for high-performance computing (HPC) applications in virtualized homogeneous clusters. To formally define the basic resource allocation problem, the authors have assumed that an application requires only one VM instance; the application's computational power and memory requirements are static and known a priori. The authors have defined a Mixed Integer Programming Model that describes the problem. The authors have proposed several heuristics to solve the problem and evaluated them experimentally across different workloads. The results show that the multi-capacity bin-packing algorithm that sorts tasks in descending order by their largest resource requirement outperforms or equals to all the other evaluated algorithms in terms of minimum and average yield, as well as failure rate. Limitations of the proposed approach are that no other system resources except for CPU are considered in the optimization and that the applications' resource needs are assumed to be known a priori, which is not typical in practice.

Cardosa et al. in [8] have investigated the problem of power-efficient VM allocation in virtualized enterprise computing environments. They leverage min, max and shares parameters, which are supported by the most modern VM managers. Min and max allow the user to specify minimum and maximum of CPU time that can be allocated to a VM. Shares parameter determines proportions, in which CPU time will be allocated to VMs sharing the same resource. Finally, the authors proposed Power Expand Min Max algorithm. In comparison to the basic policy, this algorithm uses the value of profit that can be gained by allocating an amount of resource to a particular VM. It leverages the ability to shrink a VM to min resource requirements when necessary and expand it when it is allowed by the spare capacity and can bring additional profit. The power consumption cost incurred by each physical server is deducted from the profit to limit the number of servers in use.

Verma et al. in [9] have investigated the problem of dynamic placement of applications in virtualized systems, while minimizing the power consumption and maintaining the service-level agreement (SLA). To address the problem the authors have proposed the power and migration cost aware application placement (pMapper) framework. The authors consider the problem as a bin packing problem with variable bin sizes and costs. The bins, items to pack and bin costs represent servers, VMs and power consumption of servers respectively. To solve the bin packing problem FFD algorithm has been adapted to work for differently sized bins with item-dependent cost functions.

Finally, the placement algorithm has been designed that optimizes the power and migration cost trade-off (pMaP). A VM is chosen to be migrated only if the revenue due to the new placement exceeds the migration cost. pMap searches the space between the old and new placements and finds a placement that minimizes the overall cost (sum of the power and migration costs). The authors have implemented the pMapper architecture with the proposed algorithms and performed extensive experiments to validate the efficiency of the approach. The experimental results show that the approach allows saving about 25% of power relatively to the Static and Load Balanced Placement algorithms.

III. THE ARCHITECTURE OF THE PROPOSED METHOD

In this paper, the proposed infrastructure is illustrated by 5 datacenters. Each data center has its own storage, memory, and CPU capacity and some dedicated VMs. According to Fig. 1, the network at the data center is assumed to have a 3-tier topology. Each VM has its own energy consumption and memory, storage, and CPU capacity. We assume that the user can properly select the required virtual machine. It means that if the virtual machine is fully provided with its required resources, SLA violation doesn't happen; this is one of the fundamental requirements of computational programs. Therefore, the Bin Packing problem was considered in which virtual machines are bins and tasks are the problem items. This study aims to use the VMs which have all the necessary resources, as well the least energy consumption. Consequently, the genetic algorithm was employed to solve this problem.

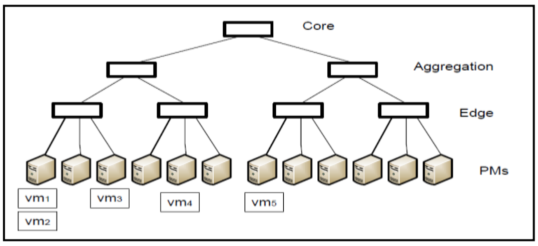

Fig.1. *The communication network of a data center [10]*

IV. THE PROPOSED GENETIC ALGORITHM

This section presents the genetic algorithm for task placement problem and the written code is presented in Fig. 2. It discusses in detail the encoding scheme, genetic operators and fitness function of the genetic algorithm (GA).

```
procedure Generate_GA_Algorithm();
 var
 co:Word;
 col,N:Byte;
 T:Chromosome2;
 Index,Index1:Byte;
begin
 for co := 1 to Ep do
 begin
  N:=GA_population_Num;
  for col := 1 to Mutation_Num do
   begin
    N:= N+1;
    Index:=Chancy_Algorithm_For_Selection;
    CHA[N]:=MultiPoint_Mutation(CHA[Index],MPoint_M);
   end;
  for col := 1 to Crossover_Num do
   begin
    Index:=Chancy_Algorithm_For_Selection;
    Index1:=Chancy_Algorithm_For_Selection;
    while index = index1 do
     Index1:=Chancy_Algorithm_For_Selection;
    T:= MultiPoint_Crossover(CHA[Index],CHA[Index1],MPoint_C);
    N:= N+2;
    CHA[N-1]:=T[1];
    CHA[N]:=T[2];
   end;
  Sort_GA(N);
  Best_Result_evry_EP[co]:=Compute_CH_power_Consume(1);
 end;
end;
```

Fig. 2. *Written code of the proposed genetic algorithm*

A. Encoding scheme

A chromosome in this GA consists of | T | genes, each of which stands for a task. The value of a gene is a positive integer between 1 and | V | representing the virtual machine where the task is allocated. Fig. 3 shows an example task placement and its corresponding chromosome.

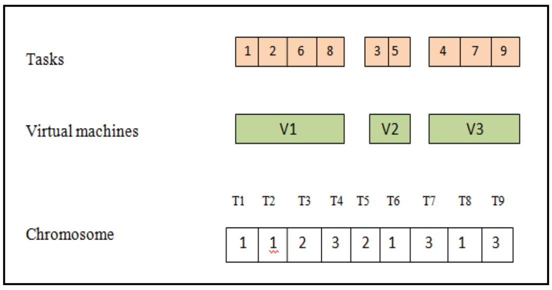

Fig. 3. *An example of task placement and its corresponding chromosome*

B. Crossover

In this paper uses multi-point crossover. More specifically, two chromosomes are randomly selected from the existing ones to perform the crossover.

C. Mutation

The mutation operator simply randomly picks up a gene in the chromosome and inverts the value of the chosen gene.

D. Fitness function

Since our goal is to reduce energy consumption, the fitness function used in this paper is a descending bubble sort function. For each allocation, it sorts virtual machines based on their energy consumption and assigns more fitness to the virtual machine having the necessary resources with the less energy consumption.

V. EVALUATION

Since the issues of energy consumption, resource allocation management and optimization in cloud computing are some important problems regarding the cloud environment, in order to evaluate the proposed method, it was executed on 5 data centers with random number of virtual machines. As Fig. 4, the evaluation of the proposed method shows that it has reduced energy consumption.

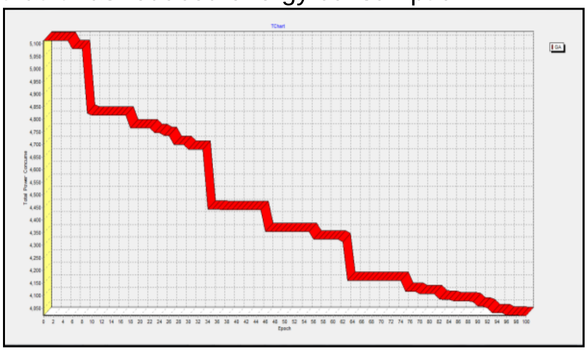

Fig. 4. *Power consumption of the system with GA*

Moreover, regarding the optimization of energy consumption in cloud computing, the comparison of several similar methods with the proposed algorithm showed that it reduces the costs. According to table 1, this method is more scalable in comparison of BFD and FFD and also has lower energy consumption.

Table 1. *Comparison of the proposed algorithm with other algorithm*

Algorithm name	Number of data centers	Number of hosts	population	Mutation and crossover	Energy consumption (watt)
FFD [10]	5	10	100	0.1	4067
BFD [3]	5	10	100	0.1	4063
GA	5	10	100	0.1	4050

VI. Conclusion

Cloud computing has attracted a lot of attention in domains of industry and education, such that it is considered the backbone of modern societies in the future. One of the challenges of this domain is the problem of energy consumption optimization, which was the target of this research and thus a solution was proposed. Clouds are usually comprised of several resources. These resources can be virtual, distributed, and heterogeneous. Since cloud computing is performed in a distributed environment, the energy consumption was evaluated at the data centers. In this paper, a genetic algorithm was proposed to efficiently allocate tasks to virtual machines, which allocates resources based on the available resources and the energy consumption of each virtual machine. This genetic algorithm was implemented, executed, and evaluated using the experiments.

Our future work focuses on optimization by combining the genetic algorithm with other algorithms, as well as considering QoS parameters in the cloud environment.

References

[1] L. Gong, J. Xie, X. Li, B. Deng, “Study on Energy Saving Strategy and Evaluation Method of Green Cloud Computing System”, in IEEE 8th Conference on Industrial Electronics and Applications (ICIEA), 978-1-4673-6322-8/13, 2013.

[2] A. Beloglazov, J. Abawajy,and R .Buyya," Energy-Aware Resource Allocation Heuristics for Efficient Managementof DataCenters for Cloud Computing", Elsevier Future Generation Computer Systems ,pp.758-761,2011 .

[3] Q. h. Nguyen, D.N. Pham, H.N. Nguyen, H.T. Nguyen, T. Nam,” A Genetic Algorithm for Power-Aware Virtual Machine Allocation in Private Cloud”, in IFIP International Federation for Information Processing, 2013.

[4] S.K. Garg , C. S. Yeo , A. Anandasivam, and R. Buyya, “Environment-conscious scheduling of HPC applications on distributed cloud-oriented data centers”, Journal of Parallel and Distributed Computing, 2010.

[5] R. Buyya, A. Beloglazov, and J. Abawajy, “Energy-Efficient management of data center resources for cloud computing: A vision, architectural elements, and open challenges”, in Proceedings of the International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA 2010), Las Vegas, USA, July 12-15, 2010.

[6] N. Kim, J.cho, E. Seo, ”Energy-Credit Scheduler: An Energy Aware Virtual Machines Scheduler For Cloud Systems”, in Elsevier Future Generation Computer Systems,2012.

[7] M. Stillwell, D. Schanzenbach, F. Vivien, and H. Casanova, “Resource allocation using virtual clusters”, in Proceedings of the 9th IEEE/ACM International Symposium on Cluster Computing and the Grid (CCGrid 2009), Shanghai, China, pp. 260–267, 2009.

[8] M. Cardosa, M. Korupolu, and A. Singh, “Shares and utilities based power consolidation in virtualized server environments”, in Proceedings of the 11th IFIP/IEEE Integrated Network Management (IM 2009), Long Island, NY, USA, 2009.

[9] A.Verma, P. Ahuja, and A. Neogi, “pMapper: power and migration cost aware application placement in virtualized systems”, in Proceedings of the 9th ACM/IFIP/USENIX International Conference on Middleware. Springer-Verlag New York, Inc. , pp. 243–264, 2008.

[10] G. Wu, M. Tang, Y. Ch. Tian, W. Li, “ Energy-Efficient Virtual Machine Placement in Data Centers by Genetic Algorithm”, in Springer-Verlag Berlin Heidelberg, pp. 315–323, 2012.

[11] P. Mell, T. Grance, "The NIST definition of cloud computing", Natl. Inst. Stand. Technol. 53 (6) (2009) 50.

[12] Ch-M. Wu, R-S. Chang, H-Y. Chan, ”A green energy-efficient scheduling algorithm using the DVFS technique for cloud datacenters”, in Elsevier Future Generation Computer Systems, Volume 37,pp 141–147, July 2014.