IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Optimized Virtual Machine Scheduling in Cloud Environment

[#]Prof. Nikul Patel, [#]Prof. Viranch Kadia

[#]Research Scholar, Department of Information Technology, Sardar Patel College of Engineering, Gujarat Technological University

Bakrol, Anand

ABSTRACT

The placement of the Virtual Machines in a Cloud Computing environment affects the application performance, which has been deployed in those virtual machines. The placement of the virtual machines considering their past, current and future requirements will lead to optimal utilization of the resources of the Cloud Service Providers (CSP). It is not just in interest of the users of the cloud services but also the CSP. The placement of the virtual machines on the appropriate hosts by the CSP will help to provide better services to the users. The requirements of the users such as required by law, for, e.g., placement of data within a certain geographic boundary or adhering to SLA(service level agreement) requirement of performance, uptime, etc. can be achieved only with proper selection of hosts. In this study, several metrics such as utilization of the CPU, memory, storage of the hosts and the I/O requirements of the applications have been considered to devise a better VM scheduling policy which can be implemented using one of the several open source hypervisors such as Xen, KVM or Virtual Box. It is also planned to deploy the policies in a real HyperBox based cloud environment.

Keywords — Cloud Computing, VM Scheduling, CSP, Quality of service.

I. INTRODUCTION

Cloud Computing is a Distributed Computing model which provides services to the customers. Cloud Providers provides services to their customers and charges as per usage by exacting customer. Customer use services when you want to use and pay for only what you have used. Cloud computing is a construct that allows you to use applications that in fact reside on a location different from your machine location. The cloud environment provides a different virtualized platform that helps user to complete their jobs with minimum completion time and minimum costs. In the cloud computing model, computing power, software, storage services, and platforms are delivered on demand to external customers in excess of the internet. Cloud makes it likely for users to use services provided by cloud providers from anyplace at any time.

Cloud services are isolated into three kinds Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS) correspondingly. The essential attributes of distributed computing, for example, asset pooling, expansive organization access, versatility, on-request benefits, physical cloud assets and middleware ability structure the premise supplier of conveying IaaS and PaaS as an assortment of straightforwardly server farms and runtime condition and piece instruments which facilitate the creation, arrangement and execution cycle of use in the cloud. At long last, to give the previously mentioned administrations, arrangement models, for example, Public Cloud,

Private Cloud, Hybrid Cloud and Community Cloud are utilized by the cloud suppliers.
The infrastructure of the cloud is provided openly to all the general public by the organization in public cloud. Anyone can access services from anywhere publicly. Where, private cloud is used for a single organization only. Community Cloud is formed by several organizations and wires a specific community that has shared concern for their upcoming use. It might be managed by the any one of the shared organization or a third party organization. Last type is Hybrid Cloud, is a cloud formed by the composition of two or more clouds that is private, community, or public. Hybrid computing is bound together by uniform technology which enables data and application portability.

VM SCHEDULING

Sceduling is an adjusting situation where cycles or assignments are arranged according to the given necessities and utilized calculation. In Cloud Computing VM planning calculations are utilized to plan the VM needs to the Physical Machines (PM) of the intense Data Center (DC) according to the necessity satisfied with the mentioned assets (for example Smash, Memory, Bandwidth and so on). In the present time there are so many cloud suppliers in market that have distinctive limit of Data Centers and Physical Machines introduced. SalesForce, Amazon, Microsoft office 365 and Windows Azure, Oracle Cloud, Google Apps and so on are significant cloud suppliers of the time of 2013. All in all booking calculation works in three levels according to given beneath :

1. For the arrangement of VMs locate the appropriate Physical Machine.
2. Decide the right provisioning plan for the VMs.
3. Scheduling the assignments on the VMs.

Chiefly planning calculations are portrayed as static or dynamic calculations. First-Come-First-Severed (FCFS) is a genuine case of static VM booking calculation. Hereditary Algorithm is a genuine case of dynamic VM booking calculation that will be talk about later in the following segment. Eucalyptus utilizes covetous or Round Robin calculation, with GREEDY the principal hub which can meet the underlying prerequisites will be picked. The OpenNebula default scheduler gives an evaluation booking strategy that places VMs on physical machines according to the evaluation of PMs.

Cloud provisioning is additionally critical figure that comes association the cloud assets in savvy, vitality proficient or secure mindful technique as examined in Introduction segment. Cloud Provisioning comprises of three stages: (1) Virtual Machine Provisioning, (2) Resource Provisioning and (3) Application Provisioning. Here, we need to consider on Virtual Machine Provisioning

II. Literature Survey

1) Effective live migration of virtual machines using partitioning and affinity aware-scheduling-2018 [1]

Anis Yazidi et.al. focused on reduce the volume of separated traffic into two categories intra-group and inter-group categories **.** One application of graph partitioning could be to group inter-communicating VMs together in such a way that the VMs with a high degree of inter-communication traffic are placed together. In this case, the VMs can be modeled as vertices, while the amount of their mutual communication corresponds to the edges in the graph. An affinity-aware algorithm has been devised to schedule the migration of VMs for the purpose of minimizing the volume of separated traffic.

2) Optimal Scheduling of VMs in Queuing Cloud Computing Systems With a Heterogeneous Workload- FEB-2018 [2]

Mian Guo et. al. discuss that the focus is on delay-optimal VM scheduling problem in a queuing cloud computing system with dynamic workload and static amount of resources. Two main algorithms have been proposed: 1. SJF-RL (Shortest Job First – Reinforcement Learning); 2. MMBF-RL (min-min Best Fit – Reinforcement Learning); in heavy loaded and highly dynamic environment

3)Self managed virtual machine scheduling in Cloud systems July2017 [3]

Stelios Sotiriadis et. al. propose VM scheduling algorithm that consider the VM resource that is already running since long time by means of analyzing the VM usage that is done past to perform VMs scheduling. The results show that our solution refines traditional instant-based physical machine selection as it learns the system behavior as well as it adapts over time. The analysis is prosperous as for the selected setting we approximately minimize performance degradation by 19% and we maximize

CPU real time by 2% when using real world workloads.

4) Intelligent Algorithms for Optimal Selection of Virtual Machine in Cloud Environment, Towards Enhance Healthcare Services-2017 [4]

Ahmed Abdelaziz et. al. evaluate genetic algorithm (GA), particle swarm optimization (PSO) and parallel particle swarm optimization (PPSO) to find optimal chosen of VMs in a cloud environment. Algorithms on MATLAB and scheduling on CloudSim conclude that PPSO algorithm is better than GA and PSO algorithms..

5) Hard Real-Time Task Scheduling in Cloud Computing Using an Adaptive Genetic Algorithm April- 2017
[5]

Amjad Mahmood et. al. proposed a greedy and adaptive genetic algorithm greedy with an adaptive selection of suitable crossover and mutation operations (named as AGA) to allocate and schedule real-time tasks with precedence constraint on heterogamous virtual machines has been proposed in the paper. The algorithm operates in an iterative manner and maintains a set of solutions, known as populations, in each iteration. It also works on the basis of historical data.

6) A Survey of Virtual Machine Placement Techniques in a Cloud Data Center DEC – 2015 [6]

Zoha Usmania et. al. present comprehensive study of the state-of-the-art VM placement and consolidation techniques used in green cloud which focus on improving the energy efficiency. A detailed comparison is presented, revealing pitfalls and suggesting improvisation methods.

7) Using Genetic Algorithm in Profile-based Assignment of Applications to Virtual Machines for Greener Data Centers-Nov-2015[7]

Meera Vasudevan et. al. proposed a Penalty-based Genetic Algorithm (GA) is presented in this paper to solve a defined profile-based application assignment problem whilst maintaining a trade-of between the power consumption performance and resource utilization performance. Case studies show that the penalty-based GA is highly scalable and provides 16% to 32% better solutions than a greedy algorithm

8)Deadline Constrained Cloud Computing Resources Scheduling for Cost Optimization Based on Dynamic Objective Genetic Algorithm - 2015[8]

Chen, Z. G. et. al. proposed a DOGA (Dynamic Objective GA) deadline-constrained workflow scheduling model on cloud computing is proposed for tight deadline condition of the jobs. Experimental results show that DOGA can find better solution with smaller cost than PSO does on different scheduling scales and different deadline conditions. DOGA approach is more applicable to be used in commercial activities.

9) A Hybrid Genetic Algorithm for Optimization of Scheduling Workflow Applications in Heterogeneous Computing Systems-2015[9]

Ahmad, S. G. et. al. discuss that efficient heuristic is incorporated in the proposed Hybrid Genetic Algorithm (HGA). A solution obtained from a heuristic is seeded in the initial population that provides a direction to reach an optimal (makespan) solution. Two fold genetic operators search rigorously and converge the algorithm at the best solution in less amount of time. The proposed algorithm also optimizes the load balancing during the execution side to utilize resources at maximum.

10) International Conference on Information and Communication Technologies (ICICT 2014) .
A Novel Family Genetic Approach for Virtual Machine Allocation [10]

Christina Terese Josepha et. al. discuss that efficiency of virtualization is a result of optimal placement of the virtual machines to suitable hosts. This paper proposes a novel technique to allocate virtual machines using the Family Gene approach. Experimental analysis proves that the proposed approach reduces energy consumption and the rate of migrations

III. Problem Statement

Existing techniques only consider the availability of resources such as CPU, memory, storage, network etc. but does not consider their speed, type, capacity, etc

It is important to highlight here that several of these parameters and metrics available from Open Stack have not been considered for scheduling, placement of VM or migration of VM by any of the previous studies.

Cloudsim identifies a Job to be allocated to a VM in the Cloud using several properties of the VM.

They have been listed in the following table

Parameter Name	Variable value (default)	Description
id	>= 0	unique ID of the VM
userId	>= 0	ID of the VM's owner
mips	> 0	the mips
numberOfPes	> 0	amount of CPUs
ram	> 0	amount of ram
bw	> 0	amount of bandwidth
Size	>= 0	amount of storage
vmm	!= null	virtual machine monitor
cloudletScheduler		cloudletScheduler policy for cloudlets

Table 1: Detail of Parameter

While these properties of a task or a VM were enough in the initial days of evolution of Cloud Computing, they do not serve the purpose with the hybrid workloads available today. The hybrid workload available today consists of several other parameters ranging from small IoT devices to large machine learning applications. For example, several cloud providers now provide hardware accelerators, graphics processing units (GPUs) and tensor processing units (TPUs). The current version of Cloudsim does not consider such requirements while scheduling tasks or the VMs over physical hosts. In the current work, several parameters have been identified from OpenStack cloud deployment software which is free and OpenSource and is widely accepted in the Cloud computing community. The current version of Cloudsim supports the following scheduling mechanisms:

Inter Quartile Range (IQR) VM allocation policy

The Local Regression (LR) VM allocation policy

The Local Regression Robust (LRR) VM allocation policy

The Median Absolute Deviation (MAD) VM allocation policy

The Static Threshold (THR) VM allocation policy

Simple VM allocation policy that dos not perform any optimization of the VM allocation.

The Maximum Correlation (MC) VM selection policy

The Minimum Migration Time (MMT) VM selection policy

The Minimum Utilization (MU) VM selection policy

The Random Selection (RS) VM selection policy

The above policies provide moderate to good results while scheduling of static VM workloads for cloud computing systems with limited hosts. Cloud Computing infrastructure consists of hundreds of thousands of physical servers and millions of virtual machines. The existing policies in Cloudsim does not scale and provide optimum results in such cases. Evolutionary algorithms such as Genetic Algorithm converge very fast and provide optimum solution in a limited amount of time. This study is based on using Genetic Algorithm with several of the parameters available with OpenStack to produce optimum scheduling result.

IV. Proposed Work and Solution

Host Parameters Identification

A host in a cloud computing system basically consist of memory (RAM), Network (I/O), Storage (I/O), Operating Systems, Hardware Accelerator and CPU (cores)

Network (I/O) :

1)Infiniband 100GBPS (Fastest) 2)40 G Ethernet (very Fast) 3)10 G Ethernet(Fast) 4)1 G Ethernet (Normal)

Operating System : Linux, Windows

Storage (I/O) :

1) DAS – Direct Attached Storage
2) NAS – Network Attached Storage
3) SSD – Solid State Drive
4) SAN – Storage And Network
5) TAPE – Tape Library

IOPS (Input Output per Second) : SAN(Fastest)<SSD<DAS<NAS<Tape(slow)

Hardware Accelerator : GPU, PCIe storage, Co-Processor

CPU Cores : 1 to 128 (or more) : NUMA (Non-Uniformed Memory Access)

Open Stack Metrics Identification

[1] CPU with values 1 t o 5

GHz → 1) 1.8 2) 2.0 3) 2.2 4) 2.4 5) 2.6

(Reference Intel Xen processor)

Other Values can also consider.

[2] Disk I/O with values 1 to 5

1) NAS 2) DAS 3) SSD 4) SAN 5) PCIe (Hardware accelerator)

[3] Regional Restriction with value 0 & 1

1)0 – Local Region

2)1- Can take resources from other data centers (other countries)
[4] Multi tenancy Restriction with values 1 to 5
1) Private Cloud 2) Public Cloud (Provider 1) 3) Public Cloud (Provider 1 & 2) 4) Public Cloud (Provider 1,2,3) 5) Public Cloud (Provider 1 to 4)
[5] Max No. Of instances supported by the cloud provider (or as configured in your account)
e.g. Account limit settings
Values 1 to 3
1) 100 2)101 to 1000 3) 1000 to 10000
[6] RAM Capacity values (for the instance type required for e.g. small/med/large)
Small: 1) 8 GB 2) 16 GB 3) 32 GB 4) 64 GB
Medium: 5) 128 GB 6) 256 GB 7) 512 GB
Large: 8)1024 GB 9)2048 GB 10) 4096 GB
It is possible to start small instances on physical host med/large RAM
[7] RAM speed with values 1 to 5
MHz : 1)1333 MHz 2) 1600 MHz 3) 2133 MHz 4) 2666 MHz 5) 3200 MHz
[8] GPU available values 1 to 5
1)1GB 2) 4 GB 3) 8 GB 4) 16 GB 5) 24 GB
(Limited to 1 VM per host GPU as GPU's are not proposed to be shared between multiple VMs)
[9] Type of host with values 1 to 3
1) x86 2) x64 3) ARM
(Default Value is: 2)
[10] Network speed with values 1 to 4
1)1 Gbps 2) 5 Gbps 3) 10 Gbps 4) 40 Gbps

Proposed Algorithm

Genetic Algorithms (GA), characterized by a binary string representation of the candidate solutions.
It can find fit solutions in a very less time. (fit solutions are solutions which are good according to the defined heuristic)
The random mutation guarantees to some extent that we see a wide range of solutions.
Coding them is really easy compared to other algorithms which does the same job.

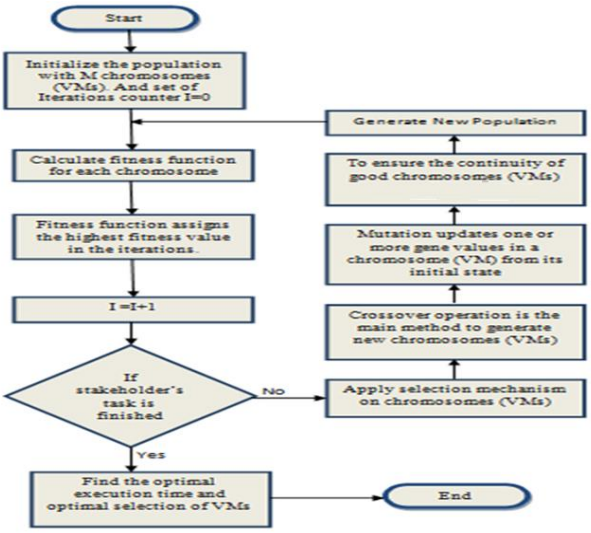

Figure 1: Proposed Genetic Algorithm

Proposed Model

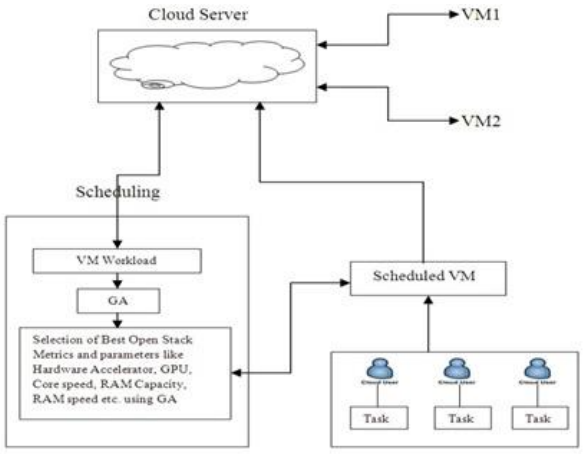

Figure 2: Proposed Architecture

Using these Open Stack Matrices and Host parameter in Genetic Algorithm and we can get better VM Scheduling result.

V. Experimental Analysis and Results

It is important to highlight here that Open stack matrices and Host parameter is used for scheduling of VM by reducing response time.

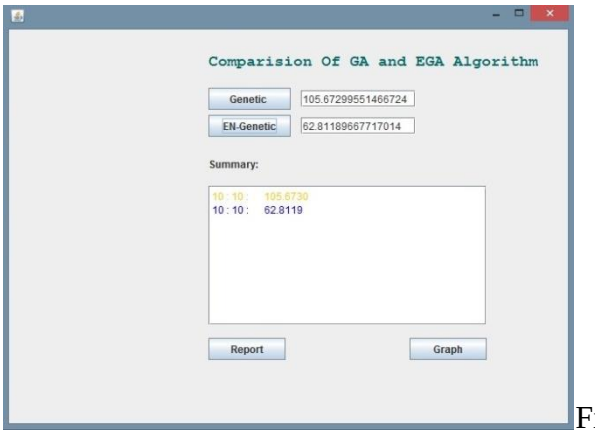

Figure 3: Makespan of 10 VMs for EGA

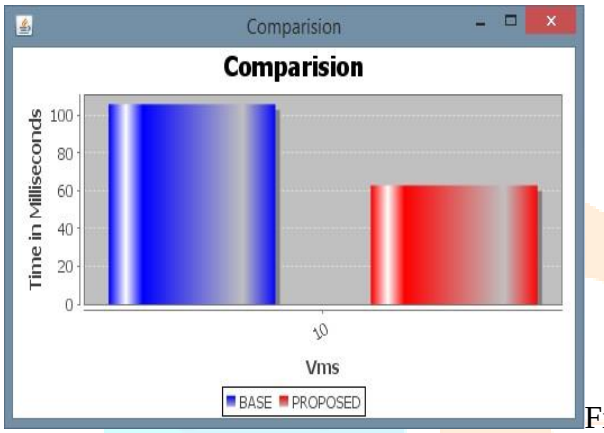

Figure 4: Comparison Makespan Graph between GA and EGA

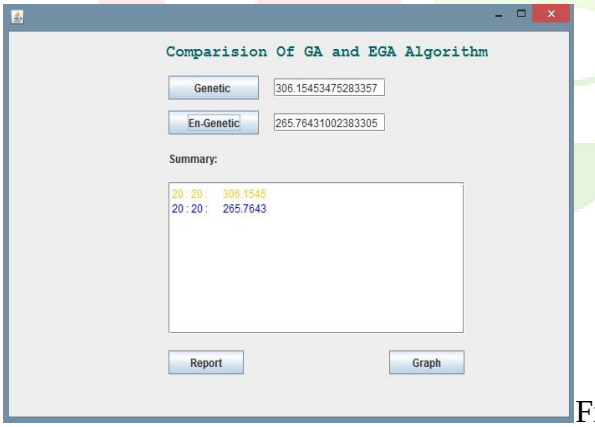

Figure 5: Makespan of 20 VMs for EGA

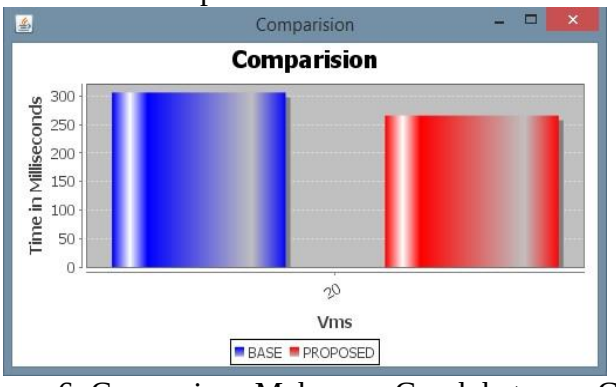

Figure 6: Comparison Makespan Graph between GA and EGA

VI. Conclusion

From the above result, it is evident that there are several ongoing efforts for proper allocation of resources and placement of virtual machines in Cloud Computing Environments. The metrics such as past, current and future requirements of the applications will be considered including the I/O requirements. The existing literature does not consider the I/O and performance requirements of the applications which this work aims to address. The problem of scheduling the VMs onto the host machines such that the number of physical hosts used is minimized, the overutilization and underutilization of the resources of a host can be identified and resolved at the same time without violating any SLA agreements. The present work will provide a metrics and heuristics based methodology for placing the VMs in the physical hosts which will minimize the number of physical hosts required for running the application. The metrics and heuristics based approach will also reduce the total energy consumption of the datacenter and high resource utilization can be achieved.

REFERENCES

[1] Yazidi, Anis, Frederik Ung, Hårek Haugerud, and Kyrre Begnum. "Effective live migration of virtual machines using partitioning and affinity aware-scheduling." Computers & Electrical Engineering 69 (2018): 240-255.

[2] Guo, Mian, Quansheng Guan, and Wende Ke. "Optimal Scheduling of VMs in Queueing Cloud Computing Systems With a Heterogeneous Workload." IEEE Access 6 (2018): 15178-15191.

[3] Sotiriadis, Stelios, Nik Bessis, and Rajkumar Buyya. "Self managed virtual machine scheduling in Cloud systems." Information Sciences 433 (2018): 381-400.

[4] Abdelaziz, Ahmed, Mohamed Elhoseny, Ahmed S. Salama, Alaa Mohamed Riad, and Aboul Ella Hassanien. "Intelligent algorithms for optimal selection of virtual machine in cloud

environment, towards enhance healthcare services." In International Conference on Advanced Intelligent Systems and Informatics, pp. 289-298. Springer, Cham, 2017.

[5] Mahmood, Amjad, and Salman A. Khan. "Hard real-time task scheduling in cloud computing using an adaptive genetic algorithm." Computers 6, no. 2 (2017): 15.

[6] Usmani, Zoha, and Shailendra Singh. "A survey of virtual machine placement techniques in a cloud data center." Procedia Computer Science 78 (2016): 491-498.

[7] Vasudevan, Meera, Yu-Chu Tian, Maolin Tang, Erhan Kozan, and Jing Gao. "Using genetic algorithm in profile-based assignment of applications to virtual machines for greener data centers." In International Conference on Neural Information Processing, pp. 182-189. Springer, Cham, 2015.

[8] Chen, Zong-Gan, Ke-Jing Du, Zhi-Hui Zhan, and Jun Zhang. "Deadline constrained cloud computing resources scheduling for cost optimization based on dynamic objective genetic algorithm." In Evolutionary Computation (CEC), 2015 IEEE Congress on, pp. 708-714. IEEE, 2015.

[9] Ahmad, Saima Gulzar, Chee Sun Liew, Ehsan Ullah Munir, Tan Fong Ang, and Samee U. Khan. "A hybrid genetic algorithm for optimization of scheduling workflow applications in heterogeneous computing systems." Journal of Parallel and Distributed Computing 87 (2016): 80-90.

[10] Joseph, Christina Terese, K. Chandrasekaran, and Robin Cyriac. "A novel family genetic approach for virtual machine allocation." Procedia Computer Science 46 (2015): 558-565.

[11] Xu, Minxian, Wenhong Tian, and Rajkumar Buyya. "A survey on load balancing algorithms for virtual machines placement in cloud computing." Concurrency and Computation: Practice and Experience 29, no. 12 (2017): e4123.

[12] Gonçalves, José Fernando, Jorge JM Mendes, and Mauricio GC Resende. "A genetic algorithm for the resource constrained multi-project scheduling problem." European Journal of Operational Research 189, no. 3 (2008): 1171-1190.

[13] Park, Byung Joo, Hyung Rim Choi, and Hyun Soo Kim. "A hybrid genetic algorithm for the job shop scheduling problems." Computers & industrial engineering 45, no. 4 (2003): 597-613.

[14] Ozdamar, Linet. "A genetic algorithm approach to a general category project scheduling problem." IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 29, no. 1 (1999): 44-59.

[15] Wang, Lee, Howard Jay Siegel, Vwani P. Roychowdhury, and Anthony A. Maciejewski. "Task matching and scheduling in heterogeneous computing environments using a genetic-algorithm-based approach." Journal of parallel and distributed computing 47, no. 1 (1997): 8-22.

[16] Hartmann, Sönke. "A competitive genetic algorithm for resource constrained project scheduling." Naval Research Logistics (NRL) 45, no. 7 (1998): 733-750.