Cost Aware Genetic Optimization for Task Allocation in Cloud Infrastructure

A Dissertation Work Submitted in
the Partial Fulfillment of
Requirement for the Award of
Degree of
Master of Technology
In
Computer Science & Engineering

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
Radharaman Institute of Technology & Science BHOPAL

Under the supervision of
Prof. Himanshu Yadav
Dr. Anurag Jain

Submitted By
Manisha Gupta
Enroll. NO. 0132CS14MT21

Contents

- Introduction
- Problem statement
- Literature Review
- Proposed Algorithm
- Pseudo Code
- Flow diagram
- Results
- Conclusion and Future Scope

What is “Cloud”

“A 'cloud' is an elastic execution environment of resources with high computation power, reliability, availability and a specified level of quality of service in distributed environment “

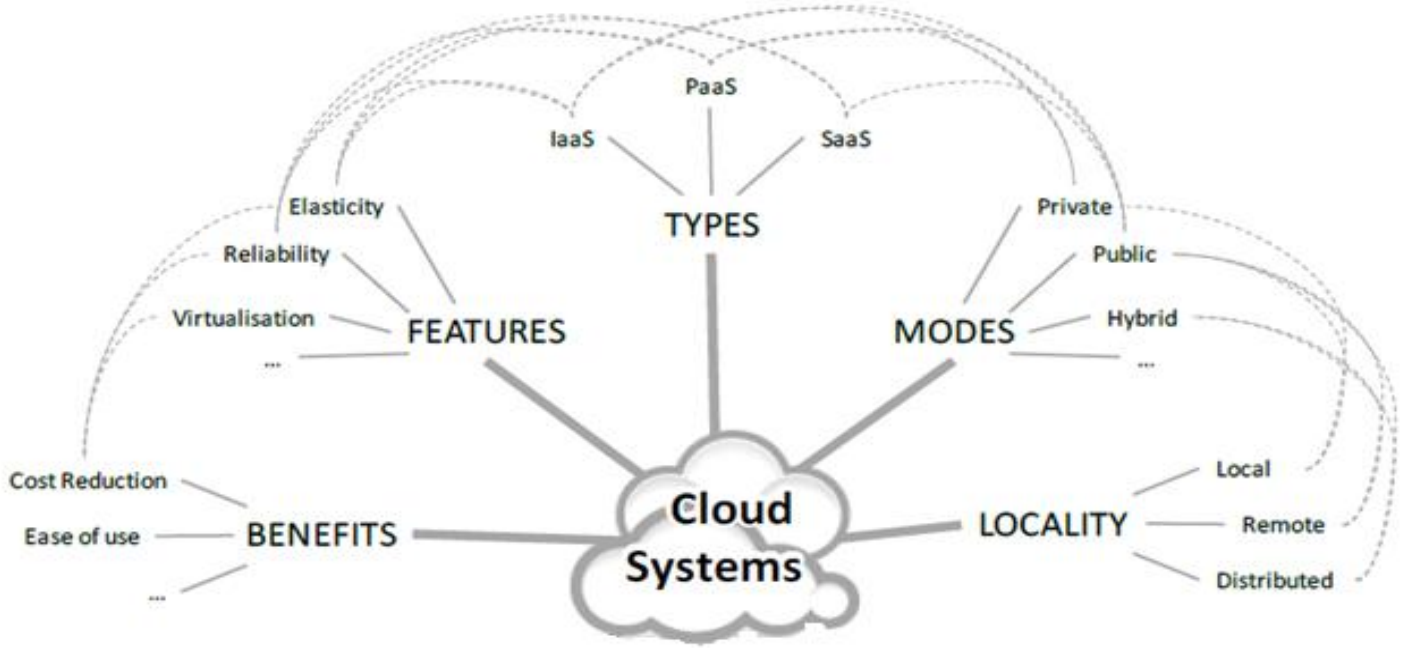


Fig 1 Cloud Systems

Characteristics

- On-demand services
- Pay-per use model
- Flexible Provisioning : Dynamically scale resources
- Networked Shared Resources
- Reliability
- Virtualization (key technology)

Motivation

- In cloud environment resource management i.e.
  - Resource selection
  - Cost minimization
  - Cloud Scalability
  - Energy efficiency
  - Load balancing are critical issue

- Several algorithm are proposed to provide efficient resource management.

PROBLEM STATEMENT

To enhance the QoS and efficiency of cloud IaaS through efficient Resource allocation .

QoS can be defined as:

- Least Cost
- Least Total Execution time
- Global Best Solution

LITERATURE REVIEW

Literature Review : Different Scheduling Algorithms

TITLE	AUTHOR(S)	YEAR	Parameters Optimized
Non pre emptive online scheduling algorithm	R Santhosh, T Ravichandran	2013	Penalty(min)
Real time scheduling using checkpointing algorithm	R Santhosh, T Ravichandran	2013	Execution time and Penalty(min)
Policy based resource allocation in IaaS cloud	Amit Nathani, Sanjay Chaudhary, Gaurav Somani	2011	Maximize utility and efficiency Minimize response time
Max min algorithm	Upendra Bhoi, Purvi M Ramanuj	2013	Execution Time (min)

Literature Review (Contd…)

TITLE	AUTHOR(S)	YEAR	Parameters Optimized
Enhanced Max-min Task Scheduling Algorithm	Upendra Bhoi, Purvi M Ramanuj	2013	Execution Time (min)
User-priority guided Min-Min scheduling algorithm	Huankai Chen, Frank Wang, Na Helian	2013	Execution time(min) with priority
Symbiosis: Network-Aware Task Scheduling in Data-Parallel F/Ws	Jingjie Jiang, Shiyao Ma, Bo Li, Baochun Li	2016	Network delay (min)
A genetic algorithm inspired task scheduling in cloud computing	Mohit Agarwal, Gur Mauj Saran Srivastava	2017	Execution time

Comparison of various Cloud Simulators

Simulator	Programming Language	Software\Hardware	Algorithm Type
CloudSim	Java	Software	Scheduling
CloudAnalyst	Java	Software	Scheduling & Load balancing
GreenCloud	C++,Otcl	Software	Network optimization
NetworkCloudsim	Java	Software	Network optimization
GroundSim	Java	Software	Network optimization
DCSim	Java	Software	Network optimization
CloudReport	Java	Software	Load balancing
SimCloud	Java	Software	Load balancing
iCanCloud	C++,Otcl	Software	Network optimization

PROPOSED WORK

Cost Based Genetic Algorithm

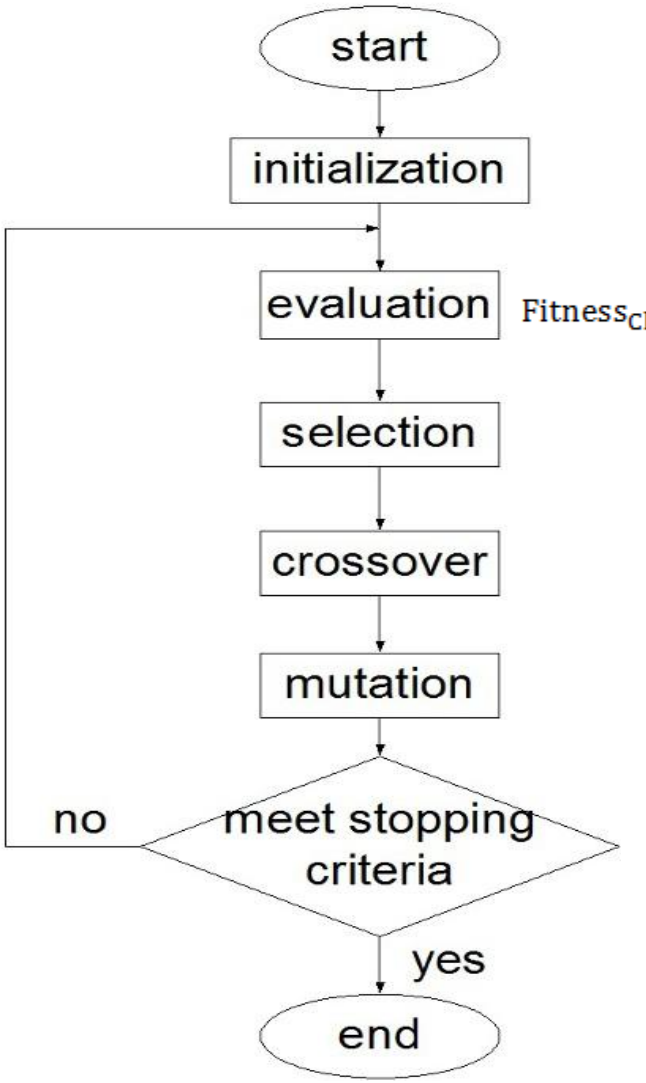
Proposed GA for task allocation is divided into 4 phases as follows:

I. Initialization

II. Evaluation and selection

III. Crossover

IV. Mutation



$$\text{Fitness}_{\text{Chromosome}_i} = \frac{1}{\alpha(\text{Total}_{\text{Cost } i}) + \beta(\text{Total}_{\text{time}_i})}$$

Figure 2: Proposed Genetic Algorithm

Parameters

- VM_MIPS i : MIPS of ith virtual machine
- T_Length i : Length of ith Task
- costPerPE : The cost of using processing in this resource
- costPerMIPS : The cost of using MIPS () in this resource
- costPerRAM : The cost of using memory in this resource
- costPerStorage : The cost of using storage in this resource
- costPerBw : The cost of using bandwidth in this resource

$$\mathrm{T}_{\downarrow}\mathrm{Exei} = \left(\frac{T_{\downarrow}Length\ i}{\mathrm{VM}_{\downarrow}MIPS\ i}\right) \quad (1)$$

$$\mathrm{Total}_{\downarrow}\mathrm{time} = \sum_{\mathrm{i}=1}^{\mathrm{n}} \frac{\mathrm{T}_{\downarrow}\mathrm{length}_{\mathrm{i}}}{\mathrm{VM}_{\downarrow}\mathrm{MIPS}_{\mathrm{i}}} \quad (2)$$

T_cost i : The total cost of executing the task

$$Cost_i = (T_{Exei_i} * (V_i.costPerRAM + V_i.costPerStorage + V_i.costPerBW + V_i.costPerPE)) \quad (3)$$

$$Total\_Cost = \sum_{i=1}^{n} Cost_i \quad (4)$$

The fitness value for a chromosome is defined by the fitness function gives as:

$$Fitness_{Chromosome_i} = \frac{1}{\alpha(Total_{Cost\ i}) + \beta(Total_{time_i})} \quad (5)$$

Where

$$\alpha + \beta = 1 \quad (6)$$

PSEUDO CODE

Algorithm: Cost Based Genetic Algorithm Task Allocation
Input: VM list VM_i, Task list T_i, Population size Po, Iteration Itr

```
1   CGATA (VM_i, T_i, Po, Itr)
2   VM_i ← VM_List ( )
3   i ← No. of VM
4   T_i ← Task_List ( )
5   C ← Genetic_algo (VM_i, T_i, Po, Itr)
6   Allocate_Resource (C) //processing the client request
```

Output: All requests have been scheduled

Algorithm: Genetic Algorithm
Input: VM list VM_i, Task list T_i, Population size Po, Iteration Itr

```
1   CGATA (VM_i, T_i, Po, Itr)
2   Po ← Initiate_Population( T_i )
3   Evaluation ( )
4   C1 ← GetFittest1 ( )
5   C2 ← GetFittest2 ( )
6   Crossover (Po, C1, C2)
7   Mutation (C1, C2)
8   Return (GetFittest ( ))
```

Output: Server with minimum fitness value

Algorithm: Evaluation

1 Evaluation ( )
2 for each C_i 0 to Po do
3 for each T_i
4 $$exec = \frac{1}{\alpha(Total_{Cost\ i}) + \beta(Total_{time_i})}$$
5 $Fitness_i = exec$
6 end for
7 end for

Output: To evaluate the fitness of all hosts

Algorithm: Allocate_Resource

Input: Chromosome List C

1 Allocate_Resource (C)
2 $C_i \leftarrow$ GetChromosomes ( )
3 for each C_i
4 Allocate (C_i)
5 end

Output: To allocate the best schedule

SIMULATION AND RESULTS

SIMULATION ENVIRONMENT

Table 2 Simulation Setup

Server	RAM(Mb)	MIPS	Storage (Gb)	core	PE	HOST
S1	2000	10000	100000	4	10	2
S2	2000	10000	100000	6	10	2
S3	2000	10000	100000	6	10	2
S4	2000	10000	100000	6	10	2
S5	2000	10000	100000	6	10	2

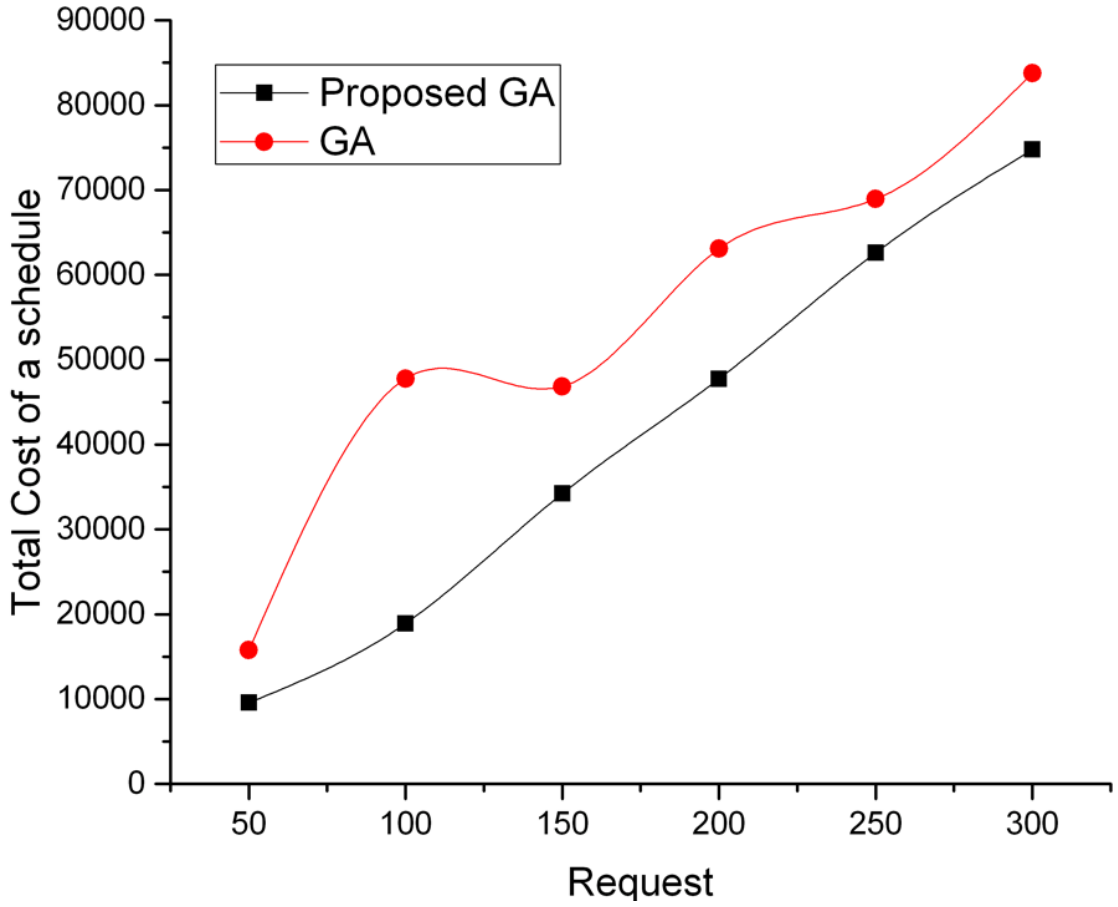
Table 3 Virtual Machine type

VM Type	Image Size	RAM	MIPS	PE	Bandwidth
VM1	10000	512	250	1	1000
VM2	10000	512	500	1	1000
VM3	10000	512	700	1	1000

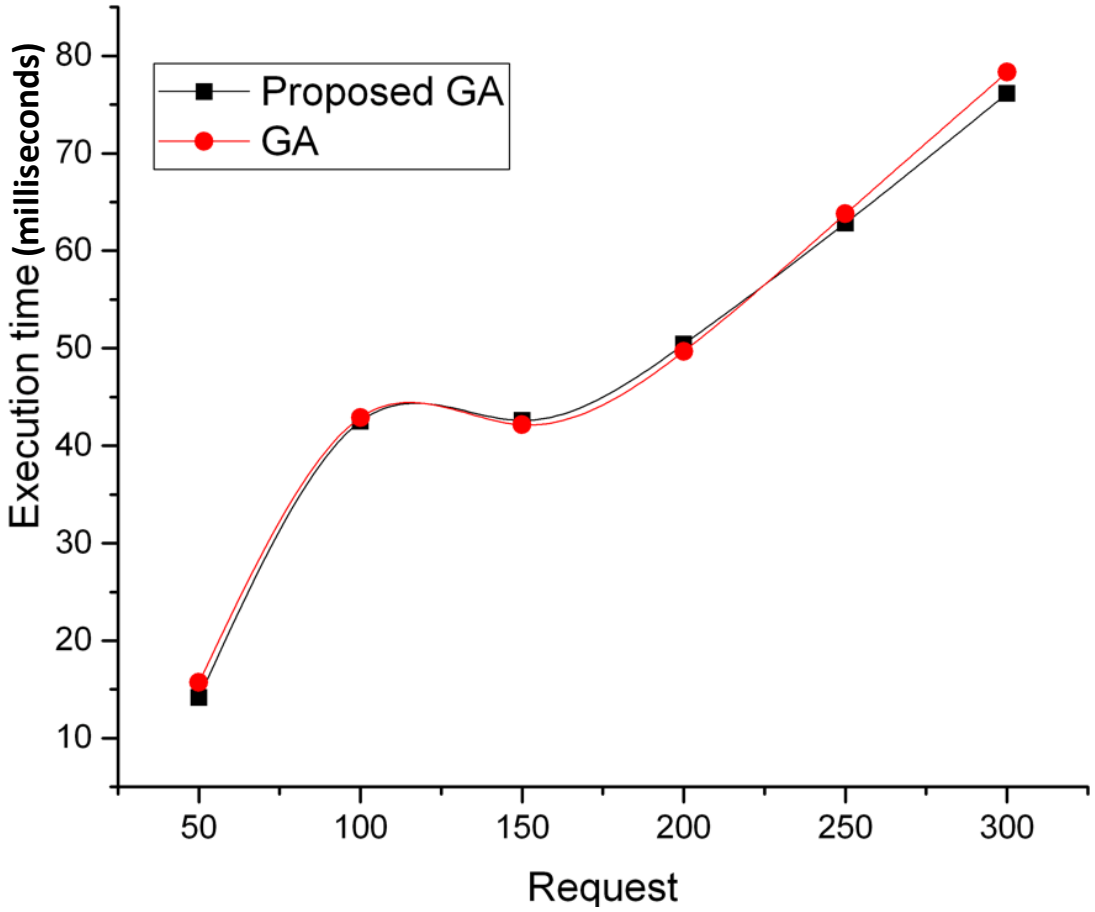
Table 4 Task Type

Task Type	Task Length	File size	Output file Size	PE
Task1	4000	300	300	1
Task2	2000	300	300	1
Task3	400	300	300	1

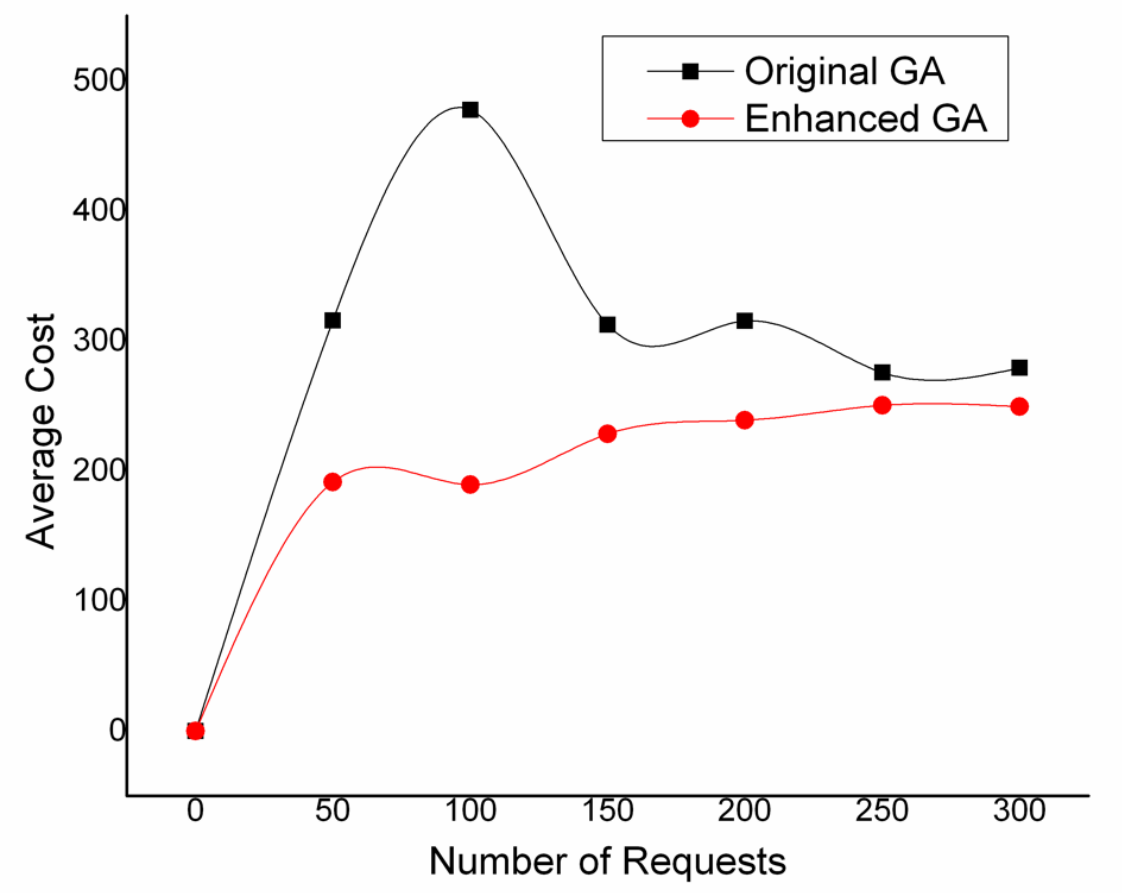
RESULTS



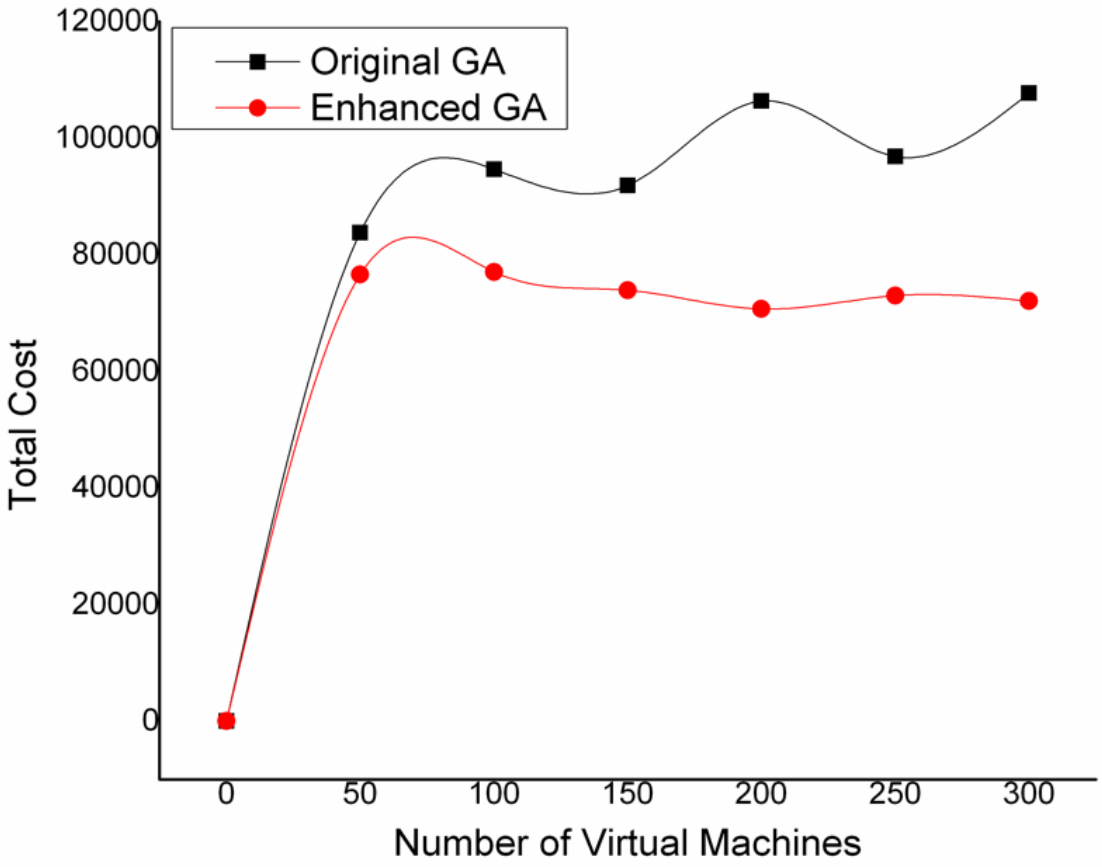
Graph 1: Comparison of cost of OGA and EGA for different no of cloudlets



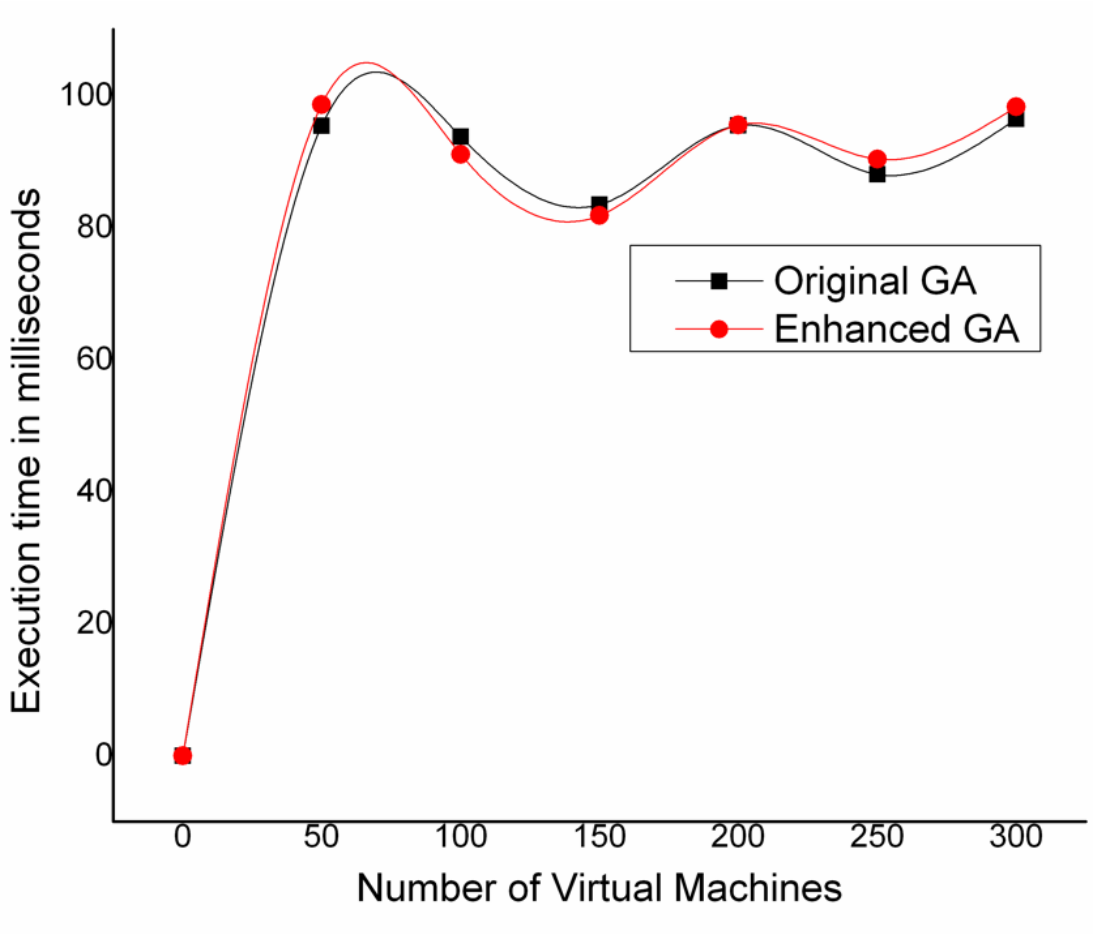
Graph 2: Execution Time (milliseconds) of OGA and EGA for different no of cloudlets



Graph 3: Average cost/cloudlet of OGA and EGA for different no of cloudlets



Graph 4: Fittest cost of OGA and EGA for different no of VMs



Graph 5: Execution time of OGA and EGA for different no of VMs

CONCLUSION

- From experimental result section, it is clear that proposed Genetic algorithm algorithm provides better QoS (Quality of service) as compared to previous proposed GA algorithm.
- Proposed algorithm completes maximum number of requests with least execution time, proposed algorithm shown that it can provide better execution time and cost over large requests with reduces average execution time over the system.

LIST OF PUBLICATIONS

- Gupta M, Jain A. “A Survey On Cost Aware Task Allocation Algorithm For Cloud Environment”. In Proceedings of the 4th International Conference on Signal Processing, Computing and Control (ISPCC 2017) 2017 (IEEE, SCOPUS) (Presented & In press).
- Gupta M, Jain A . Cost Aware Genetic Optimization for Task Allocation in Cloud Infrastructure. International Journal of Pure and Applied Mathematics (IJPAM), 2017, 117(16), pp: 613-620. (SCOPUS).

REFERENCES

1. B. Richard. "Report to congress on server and data center energy efficiency: Public law 109-431.", Lawrence Berkeley National Laboratory , pp:0-10, 2008.
2. Nathani, A., Chaudhary, S. "Policy based resource allocation in IaaS cloud." Future Generation Computer Systems28.1, 2012, pp: 94-103.
3. Guan, H., Qi, Z., Hou, Y. and Gao, Y..”A multi-objective ant colony system algorithm for virtual machine placement in cloud computing”. Journal of Computer and System Sciences, 79(8), pp.: 1230-1242.
4. Dasgupta, Kousik, et al.. “A genetic algorithm (GA) based load balancing strategy for cloud computing”. Procedia Technology, 2013, 1(10), PP: 340-347.
5. P.D., Nam, Quang-Hung, & N., Nien,. “A genetic algorithm for power-aware virtual machine allocation in private cloud”. In Information and Communication Technology, 2013, pp. 183-191.
6. Jena RK. “Multi Objective Task Scheduling in Cloud Environment Using Nested PSO Framework". Procedia Computer Science,2015 , 57, PP:1219-1227.
7. Abrishami, S. and Naghibzadeh, M .”Deadline-constrained workflow scheduling in software as a service cloud”. Scientia Iranica, 2012, 19(3),pp: 680-689.
8. Bhoi, U. and Ramanuj, P.N ,“Enhanced max-min task scheduling algorithm in cloud computing”. International Journal of Application or Innovation in Engineering and Management, 2013, vol:2(4), pp:259-264.
9. Gupta, P., S. P. Ghrera, “Load and Fault Aware Honey Bee Scheduling Algorithm for Cloud Infrastructure”. In Proceedings of the 3rd International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA) 2014,pp: 135-143.
10. Komaki GM, Kayvanfar V , “Grey Wolf Optimizer algorithm for the two-stage assembly flow shop scheduling problem with release time”. Journal of Computational Science,2015 , 8, pp:109-120.
11. Wickremasinghe B, Calheiros RN, Buyya R. “ Cloudanalyst: A cloudsim-based visual modeller for analysing cloud computing environments and applications ”. InAdvanced Information Networking and Applications (AINA), 2010 24th IEEE International Conference on 2010 Apr 20 , pp: 446-452. .
12. Calheiros, R. N., Ranjan, R., Beloglazov, A., De Rose, C. A., & Buyya, R. (2011). CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms. Software: Practice and Experience, 41(1), 23-50.
13. Suraj SR, Natchadalingam R, "Adaptive Genetic Algorithm for Efficient Resource Management in Cloud Computing." International Journal of Emerging Technology and Advanced Engineering 2014, pp: 350-356.
14. Santhosh R, Ravichandran T. “Non-Preemptive Real Time Scheduling using Checkpointing Algorithm for Cloud Computing”. International Journal of Computer Applications. 2013 Jan 1;80(9), pp: 1-27.
15. Nathani A, Chaudhary S, Somani G. “Policy based resource allocation in IaaS cloud” . Future Generation Computer Systems. 2012 Jan 31;28(1),pp: 94-103.

16. R., Buyya, R. Calheiros, R. N &, Ranjan, ,“CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms”. Software: Practice and Experience”,2011, vol:41(1), PP: 23-50.

17. Bhoi U, Ramanuj PN. “ Enhanced max-min task scheduling algorithm in cloud computing”. International Journal of Application or Innovation in Engineering and Management. 2013 Apr;2(4), pp:259-64.

18. Chen H, Wang F, Helian N, Akanmu G. “User-priority guided Min-Min scheduling algorithm for load balancing in cloud computing”. InParallel Computing Technologies (PARCOMPTECH), 2013 National Conference on 2013 Feb 21, pp. 1-8.

19. Jiang J, Ma S, Li B, Li B. “Symbiosis: Network-aware task scheduling in data-parallel frameworks”. InComputer Communications, IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on 2016 Apr 10, pp. 1-19.

20. Liu Z, Wang S, Sun Q, Zou H, Yang F. “Cost-aware cloud service request scheduling for SaaS providers”. The Computer Journal. 2013 Feb 5,pp: 1-9.

21. Coutinho, Rafaelli de C., Lúcia MA Drummond, Yuri Frota, and Daniel de Oliveira. "Optimizing virtual machine allocation for parallel scientific workflows in federated clouds." Future Generation Computer Systems 46 (2015), pp: 51-68.

22. Alkhanak, Ehab Nabiel, Sai Peck Lee, and Saif Ur Rehman Khan. "Cost-aware challenges for workflow scheduling approaches in cloud computing environments: Taxonomy and opportunities." Future Generation Computer Systems 50 (2015), pp: 3-21.

23. Ding, Youwei, Xiaolin Qin, Liang Liu, and Taochun Wang. "Energy efficient scheduling of virtual machines in cloud with deadline constraint." Future Generation Computer Systems 50 (2015), pp: 62-74.

24. Lu, Kuan, Ramin Yahyapour, Philipp Wieder, Edwin Yaqub, Monir Abdullah, Bernd Schloer, and Constantinos Kotsokalis. "Fault-tolerant Service Level Agreement lifecycle management in clouds using actor system." Future Generation Computer Systems 54 (2016), pp: 247-259.

25. Hussain, Hameed, Saif Ur Rehman Malik, Abdul Hameed, Samee Ullah Khan, Gage Bickler, Nasro Min-Allah, Muhammad Bilal Qureshi et al. "A survey on resource allocation in high performance distributed computing systems." Parallel Computing 39, no. 11 (2013), pp: 709-736.

26. Gao, Yongqiang, Haibing Guan, Zhengwei Qi, Yang Hou, and Liang Liu. "A multi-objective ant colony system algorithm for virtual machine placement in cloud computing." Journal of Computer and System Sciences 79, no. 8 (2013), pp: 1230-1242.

27. Tsai, Jinn-Tsong, Jia-Cen Fang, and Jyh-Horng Chou. "Optimized task scheduling and resource allocation on cloud computing environment using improved differential evolution algorithm." Computers & Operations Research 40, no. 12 (2013), pp: 3045-3055.

28. Manvi, Sunilkumar S., and Gopal Krishna Shyam. "Resource management for Infrastructure as a Service (IaaS) in cloud computing: A survey." Journal of Network and Computer Applications 41 (2014), pp: 424-440.

29. Gai, Keke, Meikang Qiu, Hui Zhao, Lixin Tao, and Ziliang Zong. "Dynamic energy-aware cloudlet-based mobile cloud computing model for green computing." Journal of Network and Computer Applications 59 (2016), pp: 46-54.

30. Javadi, Bahman, Jemal Abawajy, and Rajkumar Buyya. "Failure-aware resource provisioning for hybrid Cloud infrastructure." Journal of parallel and distributed computing 72, no. 10 (2012), pp: 1318-1331.

31. Cheraghlou, Mehdi Nazari, Ahmad Khadem-Zadeh, and Majid Haghparast. "A survey of fault tolerance architecture in cloud computing." Journal of Network and Computer Applications 61 (2016), pp: 81-92.

32. Lin, Chun-Cheng, and Der-Jiunn Deng. "Dynamic load balancing in Cloud-based multimedia system using genetic algorithm." In Advances in Intelligent Systems and Applications-Volume 1, Springer Berlin Heidelberg, 2013, pp: 461-470.

33. Mao, Yingchi, Xi Chen, and Xiaofang Li. "Max–min task scheduling algorithm for load balance in cloud computing." In Proceedings of International Conference on Computer Science and Information Technology, Springer India, 2014, pp: 457-465.

34. Sun, Weifeng, Ning Zhang, Haotian Wang, Wenjuan Yin, and Tie Qiu. "PACO: A period ACO based scheduling algorithm in cloud computing." In Cloud Computing and Big Data (CloudCom-Asia), 2013 International Conference on, IEEE, 2013, pp: 482-486.

35. Feller, Eugen, Louis Rilling, and Christine Morin. "Energy-aware ant colony based workload placement in clouds." In Proceedings of the 2011 IEEE/ACM 12th International Conference on Grid Computing, IEEE Computer Society, 2011, pp: 26-33.

36. Dutta, D., and R. C. Joshi. "A genetic: algorithm approach to cost-based multi-QoS job scheduling in cloud computing environment." In Proceedings of the International Conference & Workshop on Emerging Trends in Technology,. ACM, 2011, pp : 422-427.

37. Pardeep, and Amandeep Verma. "Scheduling using improved genetic algorithm in cloud computing for independent tasks." In Proceedings of the International Conference on Advances in Computing, Communications and Informatics, ACM, 2012, pp: 137-142.

38. Agarwal M, Srivastava GM. “A genetic algorithm inspired task scheduling in cloud computing”. InComputing, Communication and Automation (ICCCA), 2016 International Conference on 2016 Apr 29, pp. 364-367.