ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 1, January 2018

Resource Provision and Allocation Algorithms In Cloud Computing: A Survey

[1] Kalaiyarasi N, [2] Dr.Madhumathi R
[1] Department of Computer Science & Engineering Sri Ramakrishna Engineering College Coimbatore, India

Abstract: **- Cloud computing is an on-demand service resource which includes applications to data centres on a pay-per-use basis. In order to provide and allocate these resources properly and satisfy users' demands, an efficient and flexible resource allocation mechanism is needed. Due to increasing user demand, the resource provision and allocating process has become more challenging and difficult. One of the main focuses of research scholars is how to develop optimal solutions for this process. In this paper, a survey on resource provision and allocation algorithms is discussed.**

Index Terms: **Cloud Computing, Resource Allocation, Resource Provision.**

I. INTRODUCTION

Cloud computing is a new engenderment technology that allows the users to share resources over any communication network by using virtualization technique. The server computer plays the major role in the clouds, as it stores all the data within itself. The data on the server can be accessed by any cloud client by using the web browser. In cloud computing different resources are provided to the users with the help of dynamic resource allocation. Resource allocation is a major part of Infrastructure-as-a-service (IaaS) model of cloud system. Resource allocation is the process of allocating resources to the users according to their needs [1].

II. PROPOSED WORK

1. AUCTION MECHANISM:

1.1 Forward / Reverse Auction mechanism:

The forward auction algorithm derived from RAP (Resource Allocation Problem) Auction which appropriate modifications which enable to avoid infinite circulations in cycles. Reverse auction has the opposite interpretation, where sources with negative surplus withdraw resources from sinks [17].

1.2 Truthful Multi-Unit Double Auction (TMDA):

TMDA is asymptotically efficient, individual rational, truthful and budget-balanced. Trading prices were calculated using double auction method and resource allocation problem was solved using Cell Membrane Optimization (CMO) technique. Bid values of Cloud Resource Consumer (CRC) and Cloud Resource Provider (CRP) updates continuously in the successive rounds of auction [2].

1.3 Combinatorial Double Auction Resource Allocation (CDARA):

Most of the current market-based resource allocation models are biased in favor of the provider over the buyer in an unregulated trading environment. Four entities such as user, broker, cloud provider and cloud market place were required to model the CDARA. A cloud market place consisted of a Cloud Information Service (CIS) and an auctioneer.

III. WORKFLOWBASED ALGORITHMS

2.1 Fault-tolerant elastic scheduling algorithm for workflow (FTESW):

For Workflow, the existing fault-tolerant scheduling algorithms cannot be used in Cloud systems. It has two key features: virtualization and elasticity. This algorithm is an offline fault-tolerant elastic scheduling algorithm for workflow.

2.2 Cost and Energy Aware Scheduling (CEAS):

To minimize the execution cost of scientic workflow and to minimize the energy consumption Cost and Energy Aware Scheduling Algorithm is used. CEAS Algorithm covers five sub-domains. First the VM selection algorithm is used to apply the cost utility concept to map tasks to their optimal virtual machine (VM) types by the sub-make span constraint. Then, two tasks merging methods are employed to reduce execution cost and energy consumption of workflow. In order to reuse the idle VM instances which have been leased, the VM reuse policy is also proposed to

reuse these idle VM's. Finally, the task slacking algorithm is utilized to save the energy of leased VM's by DVFS technique.

2.3 BAT Algorithm:

BAT algorithm was developed by Xin-She Zang. For scheduling workflow tasks on cloud resources, a new meta-heuristic method called the BAT algorithm is applied. It is called so because its working is based on the echolocation behaviour of the virtual bats. This behaviour of echolocation can be formulated in a way that it associates with the objective function which is to be optimized. This algorithm is to schedule workflow tasks on cloud resources in such a way that the execution time is minimized and reliability is maximized while ensuring that the cost is within the user specified budget[3].

III. ALGORITHMS FOR DISTRIBUTE SERVICES TO VM:

3.1Profit driven scheduling:

3.1.1 Maximum Profit Algorithm:

MaxProfit assigns a service to a specific VM only if the assignment of this service yields some additional profit. As changes happen frequently in a dynamic environment, MaxProfit intends to run quickly, to allocate VM instances for a service so as to optimize profit in a greedy manner. In addition, MaxProfit handles users services in a ''first-come first-served'' manner. The algorithm calculates the additional profit that could occur. Through the assignment of the current request on each VM. It then selects the ''best'' VM based on the maximum resulting profit if not it moves to the next available VM.

3.1.2 Maximum Utilization Algorithm:

MaxUtil, is an approach for assigning service requests to instances, i.e., to maximize the utilization of the available service instances. In this way, a service provider could increase its profit by reducing costs regarding to the creation and preparation of needed resources. The goal of this algorithm is to select the VM with the lowest utilization by taking into account the profit. The arriving service requests are processed based on their arrival time.

3.1.3 Minimum Delay Algorithm:

MinDelay Algorithm looks at the problem from both the provider's and the users' point of view. This algorithm caters to the users' need for high QoS while still taking the provider's profit into account. Here, the author designed the algorithm that selects the service instance which will yield minimum processing delays for users while incorporating the provider's profit gain in its decisions [4].

IV. ENERGY EFFICIENCY ALGORITHMS

4.1 Energy-aware Task Scheduling Algorithm (EATS):

The author, Leila Ismail proposed energy-aware tasks scheduling (EATS) model, which divides and schedules a big data in the Cloud. The main goal of EATS is to increase the application efficiency and reduce the energy consumption of the underlying resources. The power consumption of a computing server was measured under different working load conditions. Experiments show that the ratio of energy consumption at peak performance compared to an idle state is 1:3. This shows that resources must be utilized correctly without scarifying performance. The results of the proposed approach are very promising and encouraging. Hence, the adoption of such strategies by the cloud provider's result in energy saving for data centers.

4.2 Optimizing Energy Efficiency In Heterogeneous Cloud:

The authors, Mohamed Abu Sharkh and Abdallah Shami, introduced a novel mathematical optimization model to resolve the problem of energy efficiency in a cloud data center. This technique depends on dynamic idleness prediction (DIP) using machine learning classifiers.

4.3 Energy Saving Task Consolidation Algorithm (ESTC):

Energy Saving Task Consolidation (ESTC) algorithm, which reduces the energy consumption by utilizing the idle time of the resources in a cloud system. ESTC achieves it by allocating few jobs to all available resources to overcome the idleness of the resources and it calculates the energy consumption on arrival of a task to make the scheduling assessment.

4.4 Power Aware Best Fit Decreasing (PABFD) scheduling Algorithm:

The strategy used in PABFD algorithm is, as the utilization is not specified by user in advance so, sorting the VM in descending order according to the time duration of VM is done first and a VM with greater time duration is allocated first and so on and under a migration policy, virtual machine is migrated from overloaded node as the upper cap of CPU utilization reaches at 85% [5].

V. RESOURCE PROVISIONALGORITHMS

5.1 Bin Packing (BP) based algorithm:

BP based algorithm views PMs as bins, VMs as items to be packed and the physical resource such as CPU, memory, bandwidth or storage as the dimensions. According to the distinctive characteristics of resources, BP and its variants are widely used in server selection phase to achieve physical node cost minimization, energy efficiency and utility maximization.

5.1.1 Issues in BP based Algorithms:

Non-fully packing, Variable item size and numerous different bins sizes, Dynamic items, Traffics between items.

5.2 Graph theory based Algorithms:

Graph theory based approaches can overcome the remedies of BP based algorithms by establishing communication between VM's and it can also depict the network infrastructure. To minimize bandwidth cost and further lessen energy caused by network related equipment's, graph partition techniques are often leveraged in a two-phase scheme which first partitions VMs and then mappings each VM cluster to different selected infrastructure node.

5.2.1 Issues in Graph theory based Algorithms:

Stochastic property, Non-Uniform size, fixed vertices weights.

5.3 Virtual Network Embedding (VNE) based Algorithms:

Virtual network embedding is a promising approach to provision resources. The request of the user is viewed as a virtual network the solution is divided into two phases: node mapping phase and link mapping phase. This approach gives considerations to both nodes clusters and links mapping. The mapping requires functionality match and capacity constraints respect. It enables optimization of node and bandwidth cost simultaneously.

5.3.1 Issues in VNE based Algorithms:

One substrate node, the optimum of the final result cannot be guaranteed.

5.4 Metaheuristic Algorithms (MT):

MH algorithms can achieve resource provision efficiency, by reducing the solution space and leveraging higher-level heuristics, such as genetic algorithm (GA), simulated annealing algorithm (SA), PSO and ant colony optimization etc. In some cases, MH can find the optimal solution. While in others, it may only return a local optimum in the space searched. Disturbance, e.g. crossover and mutation in GA, restart in SA, is usually leveraged to avoid being locked in the local optimum [6].

VI. GAME THEORETICAL APPROACH TO RESOURCE ALLOCATION (RA):

6.1 Achieving Nash Equilibrium in Cloud Resource Allocation Games (CRAGs):

A resource allocation is at Nash equilibrium if no client can decrease its cost by unilaterally changing its resource allocation, i.e., no client has any incentive to change its current strategy.

6.2 Stackelberg Equilibrium in Cloud Resource Allocation Games (CRAGs):

To overcome the sub-optimality of the Nash equilibrium Stackelberg variant is considered. Here the cloud provider imposes restrictions such that the cost of the resulting Nash assignment is close to the optimal. In this section, we first formulate a formal definition for a Stackelberg equilibrium in a CRAG and then provide two strategies, Aloof and Least Cost First, that attempt to ensure that the cost of the equilibrium reached in the Stackelberg game is close to the optimal [7][8].

VII. STATIC GREEDY ALGORITHM:

All user requests are known at the beginning of the scheduling process in Static Greedy Algorithm. Based on Budget or Deadline the most profitable schedule is obtained by sorting technique, Static Greedy algorithm is not possible in reality as all the future requests are not known [9].

VIII. MODIFIED ROUND ROBIN ALGORITHM:

This algorithm begins with the time equals to the time of first request, which changes after the end of first request. When a new request is added into the ready queue in order to be granted, the algorithm calculates the average of sum of the times of requests found in the ready queue including the new arrival request. This needs two registers as SR (sum) and AR (average) [10].

IX. SKEWNESS ALGORITHM

Skewness is used to quantify the unevenness in utilization of multiple resources on the server. By minimizing the skewness leads to combine of different combine different workloads and improve utilization of server. Load prediction, hot spot migration, and green computing are the three types of skewness keys [11].

X. VECTOR DOT ALGORITHM:

In vector dot scheduling HARMONY is used to virtualize the system. HARMONY extracts an end-to-end view of the SAN including performance and usage characteristics.

Optimize the utilization of resource includes physical servers, data centre network bandwidth and I/O bandwidth. Instead of virtual machine migration virtual storage migration is done. Extended vector product (EVP) is used to measure current utilization of resource [12].

XI. GREEN SCHEDULING ALGORITHM:

The running state of the server is determined by the green scheduling algorithm. It will turn on and turnoff servers based on load and virtual machine is allocated. Server must be in four states: OFF, ON, SHUTTING, RUNNING. Based on platform any of the state is triggered.

XII. BENCHMARK ALGORITHM:

The performance of different heuristic resource allocation algorithms is compared with a set of standards in benchmark algorithm. It can perform based on CPU utilization, which monitor the utilization threshold and VM migration .CPU utilization is based on Mean absolute deviation (MAD) and inter-quartile range [13].

XIII. CONTROL ALGORITHM:

In the control algorithm a different set of technique is used to predict non-stationary workloads of the system. In this two set of process are used Markov Host Overload Detection (MHOD) and Optimal Markov Host Overload Detection (MHOD-OPT) [14].

IV. CONCLUSION

Strategies discussed above mainly focus on Profit, Energy and auction mechanism but are lacking in some factors. Hence the survey paper will hopefully enhance future researchers to overcome the issues with secured optimal resource allocation algorithms and framework to robust the cloud computing paradigm.

V. ACKNOWLEDGEMENT

The paper is based on the work supported by the Management, Director, Principal and Head of the Department of Sri Ramakrishna Engineering College, Coimbatore. Thank you in advance for the valuable suggestions.

REFERENCES

[1] Samah Alnajdi, Maram Dogan, Ebtesam Al-Qahtani "A Survey of Resource Allocation in Cloud Computing", Science Direct, vol. 6, No. 5, 2016.

[2] Siva Theja Maguluri , R. Srikant, Lei Ying, "Heavy traffic optimal resource allocation algorithms for cloud Computing clusters", Science Direct, vol. 81, pp. 20–39, 2014.

[3] B. Magesh Kumar, C. Ramesh, "Vibrant Resource Allocation Algorithms using Virtual Machine in Cloud-Survey", Science Direct, vol.2, Special Issue 1, 2014.

[4] Pandaba Pradhan, Prafulla Ku. Behera, B N B Ray, "Modified Round Robin Algorithm for Resource Allocation In Cloud Computing", Science Direct, vol85, pp. 878 – 890, 2016.

[5] M. Miller, "Cloud Computing: Web-based applications that change the way you work and collaborate online", Que, 2008.

[6] R.Rajkumar, C.Lee, J.P.Lehoczky, and D.P.Siewiorek, "Practical solutions for QoS-based resource allocation problems". In IEEE Real-Time, 2015.

[7] Aman kumar, Emmanueel S.Pilli and R.C.Jshi," An efficient framework for resource allocation in cloud computing", in IEEE 4th ICCCNT, 2013.

[8] Chenn-Jung Huang, Chih-TaiGuan, Heng- MingChen, Yu-WuWang,Shun-ChihChang, Ching-Yu Li and Chuan HsiangWeng, "An Adaptive Resourc ManagementSchemeinCloud Computing", vol. 26, pp. 382-389, Science Direct, 2016 .

[9] X. Z. Hai Zhong, Kun Tao, "An approach to optimized resource scheduling algorithm for open-source cloud systems," The Fifth Annual China Grid Conference, 2010.

[10] M. c. D. Pandit and N. Chaki, "Resource allocation in cloud computing using simulated annealing," IEEE applications and innovations in mobile computing, 2014.

[11] E. G. Coffman, M. R. Garey, and D. S. Johnson, "Approximation algorithms for bin packing: A survey, Approximation algorithms" PWS Publishing Company 2014.

[12] K.A. Dowsland, Simulated Annealing. In Modern Heuristic Techniques for Combinatorial Problems (ed. Reeves, C.R.), McGraw-Hill, 2015.

[13] R. E. J. Kennedy, "Particle swarm optimization (pso)," vol. 96, pp. 121, IEEE International conference on Neural Networks, 2011.

[14] K. G. Thamarai Selvi Somasundaram, "Cloudrb: A framework for scheduling and managing high-performance computing (hpc) applications in science cloud," Future Generation Computer Systems, vol. 34, pp. 47–65, 2014.

[15] Chandrashekar S.Pawar and Rajnikant B.Wagh,"PriorityBasedDynamic Resource Allocation in Cloud Computing, International Symposium on Cloud and Services Computing, pp. 1-6, 2014.

[16] Jiayin Li, Meikang Qiu, Jian-Wei Niu, Yu Chen, Zhong Ming, "Adaptive Resource Allocation for Preemptable Jobs in Cloud Systems", pp. 31-36, 10th International Conference on Intelligent System Design and Application, 2015.

[17] Q. Bai, "Analysis of particle swarm optimization algorithm," vol. 3, Computer and Information Science, 2015.