International Journal of Engineering &Technology, 7 (4) (2018) 6377-6380

SPC

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET
doi: 10.14419/ijet.v7i4.16142
Research paper

Literature review and research issues: green aware cloud load balancing and scheduling techniques

Rajkumar Kalimuthu [1] *, Brindha Thomas [2]

[1] Department of Computer Science Engineering, Noorul Islam Centre for Higher Education, India
[2] Department of Information Technology, Noorul Islam Centre for Higher Education, India
**Corresponding author E-mail: rajkumarengg2020@gmail.com*

Abstract

The purpose of this paper is to present exploratory research of cloud federation service for the betterment of demand on basis of resource allocation, security and global environmental aware controls to review the literature, reports and research issues on tools/techniques/ methodologies. The current research paper highlights various definition, tools, techniques and methodologies in various researches and industries. This paper has been studied and adapted the researcher's scope, contribution and methodologies in cloud computing. Cloud computing is a new emerging technology to sharing of resource based on economic scale to achieve rationality. A cloud service can able access anywhere- any time in any of cost. In the provision of cloud service facing lot challenges to accomplish the user needs. Day by day number user access has increased so the cloud service provider (CSP) facing difficulty to deal the services in between server and client. The load balancing and scheduling techniques plays the major role of service management and cloud service provide want to achieve the goal of Quality of Service (QOS). Load balancer and scheduler are dynamically allocating and reallocating the task to respective sever with help of virtual machines. Sometimes the technology has imbalanced the services because of overload, duplication, automatic robot activities. It may lead the poor management service some cloud service provider are over utilized/underutilized, the consumption of fuel and emission of carbon also very high. In this review various techniques and algorithm are proposed to the load balancing and scheduling in Green Aware cloud system.

Keywords: *Green Aware Cloud, Load balancing, scheduling, Quality of services, Virtual machine.*

1. Introduction

The cloud computing is a network based resource shared pool model. The economic scale based models are usually pay - use basis cost model as delineated by NIST. Virtually distributed cloud computing administration categorized as, Infrastructure as a service (IaaS), Platform as a service (PaaS) and Software as a service (SaaS) [1]. Cloud computing supports an unparalleled computing ecosystem wherever provider established elastic cloud connection [2]. The services are both scientific and business-based paradigm [1]. In the green aware system were adapted and provide the services to user and owner. The characteristics of green aware cloud system (e.g.: low carbon emission, less engine temperature, budget, fuel consumption, etc). Current survey of application demand on the proficient utilization of assets in cloud environment and reduce the delay of VM migration with central process unit. The virtual machine context the cloud computing systems. The load balancing concept play the important role of number of virtual machines migrated and handling the process. If number of user request has increased the cloud environment it need an effective resource supply between the virtual machine and central process unit to reduce the delay of migration and effective processing. Current provisioning technique prediction error rate less than 15% [13]. In this case the system needed less error rate provision technique and the positioning of Virtual machine in server migration process it will lead high energy consumption by large-scale cloud data centers increases the fuel usage and damages the environment by excessive emission of carbon [15]. In the cloud computing resource management system handle the in and out process that require efficient scaling techniques. The optimal utilization of resource it help to achieve the low high throughput, less delay, carbon emission, less engine temperature, budget, fuel consumption, etc.

2. Related works

The cloud Federated environment running on hundred servers in each data centers. When handling of user task it will be handover the task via datacenter to server channel that situation facing more challenges, such as

1) Server Unavailability
2) Server overload
3) Energy consumption
4) Server and Virtual Machine (VM) migration
5) CPU provisioning

The above stated challenges need an effective load balancing technique for green aware cloud eco system to handle the user task. In the survey of early research work the following algorithms, models and techniques are practiced for load balancing and scheduling.

2.1. Contract based model

The model allocated the time slot to indicated the task stages such as T^{Begin} (Task begins), T^{End} (Task End) and all task comes to a

Copyright © 2018 Rajinikanth E, Jayashri S. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Cloud Service Provider (CSP) enter into First in first out (FIFO) queue and calculate the contract task timing based on task timing and queue time and proportional to the estimated resultant time [3].

2.2. Balanced and file reuse-replication scheduling (BARRS) algorithm

This approach produces the workflow estimation table, the table contain monetary costs and execution times. The execution tradeoff analysis model by exponential graph. The exponential graph to design the scheduling estimation trade off. The BARRS behave as a brute force technique because it produces the all possible configuration work flow estimation in table [4].

2.3. Heuristic algorithm

This algorithm includes both cyclic and acyclic based task model. The task are iterative without searching entire space, heuristic algorithm mainly proposing for graph based task computation [5].

2.4. Data aware multi-workflow scheduling

Most of the workflow model does not picture the information about which types of data handled in and out. Proposed data provide multi-workflow scheduling model determine the files and execution characteristics of I/O read and write with location file sharing to avoid the replication data in sequential and parallel multi-flow data transmission which adapted the mechanism of Bio information workflow pattern[22] [6].

2.5. Direct acyclic graph (DAG)

DAG-structured workflows over the Cloud. The dependency and parallelism embedded in a workflow requires that the tasks be dispatched to a group of distributed VMs to maximize the execution efficiency. DAG determine the best direction in the cloud specific task to reduce the energy utilizations [7].

2.6. Resource co-allocation method

The approaches following four steps of actions Steps-1: Meta services preprocessing and mapped by the high performance computation application (HPC), the preprocessing techniques are used to evaluate the start and processing time Steps-2: Resource usage monitoring is used to monitor the availability and processing of virtual machine (VM) Steps-3: Resource allocation is applied to particular cloud server Steps-4: Global resource co allocation policies are implemented to dynamically optimist the meta service execution [8].

2.7. Virtual machine scheduler

This scheduler is able to process the Input and output bound services and reduce the energy consumption as well avoid the Service level agreement violation. The proposed scheduling algorithm must balancing energy and performance in homogeneous network [9] with help of resource reorganization by concentration method to prevent the boot of new server to flow the other virtual machine (VM).

2.8. Extended intelligent water drops cloud algorithms (IWDC)

This approach is applied to optimize the scheduling of different workflow. The three phases of process are: 1: Initialization of parameters – initialize the all static parameter 2: Paths construction 3: Task assignment phase – selection based on the best Virtual machine [10].

2.9. Energy aware scaling algorithm

This approach is possible to control within the arise bound, here applies some polices to migrate the server depends on the level of regime the server changing the mode of process such as sleep state, running state, shutdown. The idle severs are to push the sleep state when the cluster increasing load and high regime gives the demand of reallocation cycle [12].

2.10. Nicble

This approach is able to anticipate the execution time with the help of CPU provision between CPU to virtual machine (VM). But Provision of error rate less than 15% [13].

2.11. Round robin algorithm

This approach assign the task to each processor in equal portion in circular manner. The resource scheduling able to handled randomly access by the virtual machine in short term time scale to find the average delay [14].

2.12. Power aware dynamic allocator

This approach manage dynamic reallocation of resources in various data center using of fuzzy controller with help of modified Dijkstra Algorithm and allocation strategies [15].

- First fit
- Best fit/worst fit
- Single/Multi objective optimization
- Joint/disjoint strategies

2.13. Hybrid method for minimizing delay

This approach of Edge cloud computing adopted mathematical model for appraising service delay, which is stated as the processing delay and transmission delay [16]. The proposed method improves the transmission and power control.

2.14. Novel gaming theory

This approach-based framework allocated resources followed by three strategies [17], such as

i) Price based – user request is driven by the price per unit
ii) Correlation based – other users accessing the same sub channel based on Bayesian model for payoff.
iii) Water Filling aims to optimize power and resource

Table 1: Literature Survey of Load Balancing and Scheduling Algorithms and Techniques

S. No.	Algorithm and Techniques	Description	Pros	Cons
1.	Contract-based Model [3]	CSPs to establish resource sharing contracts with unique datacenters apriori for fixed time intervals during a 24-hour time period.	Achieving both global resource allocation efficiency and Local goods in the profit earned.	Local operating cost is greater than some of the negotiated price in the contracts
2.	Balanced and file Reuse-Replication Scheduling (BaRRS) algorithm [4]	BaRRS splits technological workflows into number of sub-workflows to balance system usage via parallelization	It also exploits data reuse and replication techniques to optimize the amount of data	Needs to be shifted among tasks at run-times
3.	Heuristic algorithm [5]	Schedule both cyclic and acyclic workspace	Data locality could eliminate some communication overhead	Not feasible since of iterations are too large in number or may not even be recog-

				nized at compile-time
4.	Data-Aware Multiworkflow Scheduling [6]	Multiworkflow includes bioinformatic and Epigenomics apply shared input and temporal data files. Typical bioinformatic workflow has input files starting at 2 GB that generates temporal files of 6 GB.	Keeping a cached version of input and temporal files	Workflows is going to be implemented in a particular node of a cluster
5.	Direct Acyclic Graphs [7]	A good way of modeling task dependency relationships	DAG operations better for global optimization	Lack relevant information on how to deal with data files
6.	Resource Co-Allocation method [8]	Is to investigate for leverage task scheduling and resource allocation over an enlarged data platform	Efficiently leverage task scheduling and resource	Lack of deployment resources
7.	Virtual Machine (VM) scheduler[9]	Scheduler considers each VM workload type (CPU or I/O-bound) to select on its allocation	Minimize the SLA violations	Face the challenge of reducing their expenses
8.	Novel mathematical optimization Model[10]	Technique to be merged in any consolidation-based energy efficiency solution	Major advantage to cloud providers in the cases when live migration of VMs is not favor due to its action on performance.	
9.	Extended Intelligent Water Drops algorithm [11]	The algorithm used in any fields to solve optimization and complex scientific problems such as travel sales man problem, code coverage, Graphing coloring, optimization routing protocol	Optimizes the scheduling of Workflows on the cloud.	Energy usage Of the resources.
10.	Energy Aware Scaling Algorithm [12]	The algorithms is to control that the largest feasible number of dynamic servers activates within the boundaries of their respective optimal operating authorities.	(a) Migrate VMs from a server activating in the undesirable-low authorities and then switch the server to a sleep state; (b) Switch an idle server to a sleep state and reactivate servers in a sleep state when the cluster load improves; (c) Migrate the VMs from an overloaded server, a server activating in the undesirable-high authorities with applications promised to increase their demands for computing in the next reallocation cycles.	Difficult to achieve the standards of service in server migration
11.	NICBLE [13]	NICBLE support CPU resource provisioning application workload	Provide the accurate time prediction for non-adaptive workload. Predication Error rate less than 15%	Difficult to make accurate prediction if process has significant dependencies.
12.	Round Robin Algorithm [14]	Round Robin Algorithm for finding delay	Which is operated on short- term time scale	Doesn't encourage the long-term scheduling and Job accessing time is not considered.
13.	Power aware dynamic allocator [15]	Power aware dynamic allocator taken account with the help of modified Dijkstra Algorithm and allocation strategies. • First fit • Best fit/worst fit • Single/Multi objective optimization • Joint/disjoint strategies	Modified version of Dijkstra Algorithm allocated the network flow reduced the consumption of 1kW (about 3.3% of the total previous consumption	Tackle the critical issue within the software defined framework.
14.	Hybrid Method for Minimizing Service Delay [16]	Mathematical model for the service retard in Edge cloud computing system	To improve the transmission delay with power control.	When adapting mathematical model in multiple cloud environment facing high complexity to accommodate.
15.	Novel game theory [17]	Novel game-theory based schemes to decide the wireless resources allocation challenges of transmit power and wireless signals.	Transmit power and signals to share the constrained wireless resource according to their requests efficiently.	-

3. Conclusion

The current research work investigated the various techniques pros and cons of load balancing and scheduling. Presenting some innovation model to overcome the issues of server delay, quality of service, effective scheduling, less energy consumption, less environmental damage, maximize the throughput, minimize the delay, low prediction error, common framework for both dependent and independent network task.

References

[1] Peter Mell, Timothy Grance. "The NIST Definition of Cloud Computing (Draft)". *NIST*. 2011. https://doi.org/10.6028/NIST.SP.800-145.

[2] Rajkumar Buyya et al... "Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility". *Future Generation Computer Systems*, (2009) https://doi.org/10.1109/CCGRID.2009.97.

[3] Jinlai Xu and Balaji Palanisamy"Optimized Contract-based Model for Resource Allocation in Federated Geo-distributed Clouds" *IEEE Transactions on Services Computing*,(2018), pp1 https://doi.org/10.1109/TSC.2018.2797910.

[4] Israel Casas, Javid Taheri et al. "A balanced scheduler with data reuse and replication for scientific workflows in cloud computing systems" *Future Generation Computer Systems*, Vol.74, (2017), pp 168-178 https://doi.org/10.1016/j.future.2015.12.005.

[5] Tao Yang and Cong Fu "Heuristic Algorithms for Scheduling Iterative Task Computations on Distributed Memory Machines" *IEEE transactions on parallel and distributed systems*, Vol.8, (1997), pp. 6 https://doi.org/10.1109/71.595579.

[6] César Acevedo et al. "A Critical Path File Location (CPFL) algorithm for data-aware multiportfolio scheduling on HPC clusters", *Future Generation Computer Systems*, Vol.74, (2017), pp 51-62 https://doi.org/10.1016/j.future.2017.04.025.

[7] Fredy Juarez et al. "Dynamic energy-aware scheduling for parallel task-based application in cloud computing", *Future Generation Computer Systems*, Vol.78, (2016), pp 257-271 https://doi.org/10.1016/j.future.2016.06.029.

[8] Wanchun Dou et al. "A Resource Co-Allocation method for load-balance scheduling over big data platforms" *Future Generation Computer Systems*, (2017),

[9] Felipe Fernandes et al. "A virtual machine scheduler based on CPU and I/O-bound features for energy-aware in high performance computing clouds" *Computers and Electrical Engineering*, Vol.56, (2016), pp 854-870 https://doi.org/10.1016/j.compeleceng.2016.09.003.

[10] Wanneng Shu et. al, "A novel energy-efficient resource allocation algorithm based on immune clonal optimization for green cloud computing" *EURASIP Journal on Wireless Communications and Networking*, (2014) https://doi.org/10.1186/1687-1499-2014-64.

[11] Shaymaa Elsherbiny et al. "An extended Intelligent Water Drops algorithm for workflow scheduling in cloud computing environment" *Egyptian Informatics Journal*, Vol.19, (2017), pp 33-55 https://doi.org/10.1016/j.eij.2017.07.001.

[12] Ashkan Paya and Dan C. Marinescu, "Energy-aware Load Balancing and Application Scaling for the Cloud Ecosystem" *IEEE Transactions on Cloud Computing*, Vol. 5, No. 1, (2017). https://doi.org/10.1109/TCC.2015.2396059.

[13] Hong-Wei Li, Yu-Sung Wu, Yi-Yung Chen, Chieh-Min Wang, and Yen-Nun Huang "Application Execution Time Prediction for Effective CPU Provisioning in Virtualization Environment*" IEEE Transactions on Parallel and Distributed Systems* No.99 (2017).

[14] Yuan Zhang" Resource Scheduling and Delay Analysis for Workflow in Wireless SmallCloud" *IEEE Transactions on Mobile Computing*, No. 99 (2017).

[15] Giuseppe Portaluri; Davide Adami; Andrea Gabbrielli; Stefano Giordano; Michele Pagano "Power Consumption-Aware Virtual Machine Placement in Cloud Data Center" *IEEE Transactions on Green Communications and Networking*, No.99 (2017). https://doi.org/10.1109/GLOCOMW.2016.7849005.

[16] Tiago Gama Rodrigues; Katsuya Suto; Hiroki Nishiyama; Nei Kato "Hybrid Method for Minimizing Service Delay in Edge Cloud Computing Through VM Migration and Transmission Power Control" *IEEE Transactions on Computers*, Vol.66, No.5 (2017). https://doi.org/10.1109/TC.2016.2620469.

[17] Xi Zhang; Qixuan Zhu "Game-Theory Based Power and Spectrum Virtualization for Optimizing Spectrum Efficiency in Mobile Cloud-Computing Wireless Networks" *IEEE Transactions on Cloud Computing*, No. 99 (2017). https://doi.org/10.1109/TCC.2017.2727044.

[18] Huangke Chen; Xiaomin Zhu; Dishan Qiu; Ling Liu; Zhihui Du "Scheduling for Workflows with Security-Sensitive Intermediate Data by Selective Tasks Duplication in Clouds" *IEEE Transactions on Parallel and Distributed Systems*, Vol.28, No.9, (2017). https://doi.org/10.1109/TPDS.2017.2678507.

[19] Xingwei Wang; Xueyi Wang; Hao Che; Keqin Li; Min Huang; Chengxi Gao "An Intelligent Economic Approach for Dynamic Resource Allocation in Cloud Services" *IEEE Transactions on Cloud Computing*, Vol.3, No.3 (2015). https://doi.org/10.1109/TCC.2015.2415776.

[20] Wanyuan Wang; Yichuan Jiang; Weiwei Wu "Multiagent-Based Resource Allocation for Energy Minimization in Cloud Computing Systems" *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, Vol.47, No.2(2017). https://doi.org/10.1109/TSMC.2016.2523910.

[21] Song Wu; Yihong Wang; Wei Luo; Sheng Di; Haibao Chen; Xiaolin Xu; Ran Zheng; Hai Jin "ACStor: Optimizing Access Performance of Virtual Disk Images in Clouds" *IEEE Transactions on Parallel and Distributed Systems* Year: 2017, Vol.28, No. 9 (2017). https://doi.org/10.1109/TPDS.2017.2675988.

[22] Jeremy Leipzig, "A review of bioinformatic pipeline frameworks", *briefings in bioinformatics*, Vol.18, No. 3 (2017). https://doi.org/10.1093/bib/bbw020.