

The Load Balance of Sitakeboge Model under Cloud Computing

Liu Xuchong^{1,a}, Zhang Dafang^{2,b}, Zhao Wei^{3,c}, Jiang Zhongxin^{4,d}

^{1,3,4}Laboratory investigation of cybercrime Hunan Police Academy, Changsha, 410038, China

²College of information Science and Engineering, Hunan University, Hunan Changsha 410082, China

^aliuxuchong@163.com, ^bdfzhang@hnu.cn, ^ciuiuy@126.com, ^dcn_wuheng@163.com

Keywords: Cloud computing; load balancing; Job Scheduling; Sitakeboge Model

Abstract. Under Sitakeboge's Stress reaction and equilibrium recovery theory to Social distribution equilibrium failure, this paper proposes a host load balancing control model to each node based on Sitakeboge model under cloud computing. Based on CloudSim simulation software to Job scheduling algorithm under Sitakeboge model and Min-Min algorithm, the result approves the first algorithm is more Effectively guarantee the node load balancing than the next.

Introduction

The load balance and computing optimization closely contacts with job scheduling in cloud computing[1-9]. The purpose of job scheduling is realize the optimal matching of tasks and resources, to deliver different jobs to the corresponding nodes under the most reasonable way. For the speed, host load and network communication time are dynamic, the job scheduling is a NP-complete problems.

Existing job scheduling algorithms appear highly targeted, such as the minimum consumption time leads the strong computing resources is been accessed frequently, while weak ones will be idle, which lead to the utilization rate of resource decline, overload and other issues.

In cloud computing, there are two different load balances, like figure 1, the first load exist virtual machine layer, which is responsible for execution task, the second appears physics layer, provide resource. There is one chain load relationship between them. Therefore, the load balance control appear new characteristic under cloud computing, it needs new technologies and algorithms.

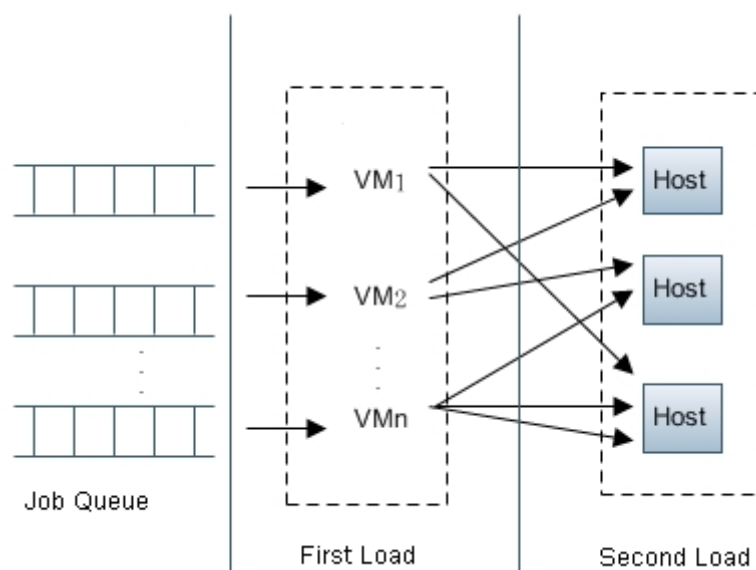


Fig 1. Cloud Computing Load

Sitakeboge et al., proposed a balance theory on a basic state of the structure, include the balance failure happen, reflect to it, and balance recovery strategy[10,12]. When the elements of state structure inconsistent with their expectation, the balance failure happen. The state structure appears

tight constraint, which guide to opposition direction. It is also instructive to the load balance control in cloud computing system, provide an idea to the load automatic adjustment mechanism.

Method

Definition

In order to introduce Sitakeboge's balance theory to cloud computing modeling, we firstly define some concepts.

Define 1. Related Element: Let $L_x = \{L_{1x}, \dots, L_{nx}\}$ means an evaluable status set to a group of load characterize $L = \{L_{1x}, \dots, L_{nx}\}$, it be used to describe the current load status; $GO_n = \{GO_{1x}, \dots, GO_{nx}\}$ is a an evaluable status set to a group load capacity target entity set $GO = \{GO_1, \dots, GO_n\}$, which descript machine load capacity. A member of $L_{ix} (i = 1, \dots, n)$ map to one of $GO_{ix} (i = 1, \dots, n)$, both are Related Elements.

Define 2. Load Status Structure(LSS): it is made of L_x and corresponding GO_x .

Define 3. Balance(B): Only if two elements of load status structure have consistent evaluation, they are balance.

Define 4. Balance Load Status Structure(BLSS): Only if all of load status structure have consistent evaluation, status structure is balance.

Define 5. Stability: Only if load status structure is balance, it is stability

Define 6. Tension: a factor to depict machine load, give a threshold, if load over it, the machine stand on unbalance status, LSS generate tightness signal and pressure, it encourage to adjust load status structure, it tend to stability structure.

Define 7. Path: if LSS is unbalanced, it generate pressure, go along the balance direction, the rule of change called path.

Stress Reaction to Balance Failure

The load status relate to current load and next load change under LSS. According to Sitakeboge theory, the change process abstract as a Markov chain. Let L_{our} be current load status, it is

$$L_{our} = \sum_{i=1}^n L_{ix}, (i = 1, \dots, n) \quad (1)$$

L_{ix} is next load under a load status of performance

$$L_{next} = L_{our} + L_{nextaddt} - L_{nextfinished} \quad (2)$$

The system comprehensive load capacity is

$$L_w = \sum_{i=1}^n GO_{ix}, (i = 1, \dots, n) \quad (3)$$

According to the definition to balance, for $i = 1, \dots, n$, LSS is balance if and only if $L_{our} < L_w$ and $L_{ix} < GO_{ix}$. For each $i = 1, \dots, n$, if $L_{ix} > GO_{ix}$, If LSS is balance failure, one or more overload nodes appears in cloud computing system. If they stay in the first layer load, it leads chain effect. If it exists in the second layer load, the load failure happen. Therefore, the balance control used in cloud computing system.

Let Load Balance(LB) be

$$LB = \frac{L_{ix}}{GO_{ix}}, (i = 1, \dots, n; 0 \leq LB \leq 1) \quad (4)$$

When $LB \geq 1$, for each $i = 1, \dots, n$, if $L_{ix} \geq GO_x$ and LSS generate tightness signal, the load be transfer under path choice rules, LSS progress toward the balance director.

Balance Recovery Strategy

Balance Recovery Strategy means the path choice rules to load transfer when balance failure exists. According to Sitakeboge theory, it set up Markov chain to load transfer.

When the balance failure exists in following time, it means bore load ($L_{nextadd}$) over the limitation of node, the load transfer is followed. For $L_{nextadd}$ has Markov property, it sets up a Markov chain to transfer.

Where, $L_{nextadd}$ assign to the i -th node ($i = 1, \dots, m$), m is host or virtual node, p_{ij} is the probability to transfer from the i -th node to j -th at the next moment, the transfer probability matrix is following:

$$p_{ij} = \begin{bmatrix} p_{00} & \dots & p_{0m} \\ \dots & \dots & \dots \\ p_{m0} & \dots & p_{mm} \end{bmatrix} \quad (5)$$

Sitakeboge et al. propose an parameter estimation solution, it also be applied in cloud computing. Each mode maintains a load balance factor---Tightness, the nodes be order as inverted sequence according to Tightness. If a node has a light tightness, it be transfer under a heavy probability. The transfer probability matrix in this analogy.

The following is to calculate the probability:

Let the current tightness of node T_j equates the current load balance degree, in other word, $T_j = LB_j$, ($0 \leq j \leq n$), $LB_j (0 \leq LB_j \leq 1)$, to calculate P_j , the system divide T_j of all nodes, except unbalance load nodes, $P_i = 1/T_j$, it is followed by normalized and map to between 0 and 1. The normalized formula is following

$$P'_j = (curP_j - \min P_j) / (\max P_j - \min P_j) \quad (6)$$

Where, P'_j is the new transfer probability from i -th node to j -th node, Eq.5 determined using P'_j . The load transfer path be decided under it.

Result and Analysis

Min-Min [13] is a widely used algorithm in task scheduling. There are many algorithms based on it. According to load balance coefficient, to choice suitable scheduling strategy, its performance be approved by theory analyze and experiment results. The Makespa of task scheduling is smaller. In here, we compare the new algorithm---cloud resource scheduling algorithm (BMB) based on stress reaction of balance failure and balance recovery theory with Min-Min algorithm.

Experiment Environment

In here, we test algorithms on CloudSin[14]. It published by Grid Laboratory of Melbourne University, Australia, and Gridbus project group used to cloud computing simulation. The main goal is to quantity and compare the performance of strategy of scheduling and allocation to different application and service model on cloud computing infrastructure, in order to control the use of cloud computing resource. Using CloudSim, the researchers and developers only focus the abstract algorithms, strategy and protocol, do not need to think of infrastructure and service and so on.

Experiment Design

In testes, every tasks are assumed to be independent, one cloud user submit 20 tasks, they be scheduled to five resources. A continue ETC(excepted Time to Compute) matrix be randomly generated. We take 10 simulations and use the average of them to analyse and compare the performance of algorithms. Table 1 simulates five different characteristic, performance and configure cloud computing resources.

Table1. Experiment resource configure

Cloudlet ID	MIPS	ProcessingCost	VMBandWith	CPUs	Memory
1	400	470	3000	2	1024
2	300	710	1000	2	1024
3	340	830	3000	2	1024
4	250	1500	1000	2	1024
5	500	2000	1200	4	2048

Experiment Result

The paper mainly analyses the equilibrium of system resource load and task scheduling length under scheduling algorithms. The figure 2 notes the comparison of different algorithms (BMB and Min-Min) on system resource load balance coefficient. Before task scheduling, each resource is free, that is to say, the initial load balance coefficient is 1. In figure 2, the load balance coefficient

still equals 0 after 11 tasks delivered under Min-Min scheduling strategy. It contains there is at least a free resource in system. Furthermore, the load balance coefficient of Min-Min less than that of BMB after scheduling. It means BMB algorithm lead the system load balance to a better status.

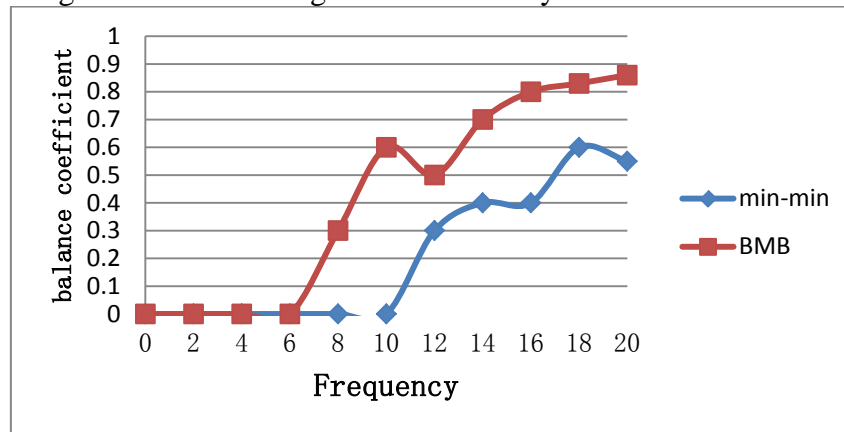


Fig. 2 The comparing of load balance coefficient

Conclusion

The paper proposes a stress reaction scheme and balance recovery strategy under Sitakeboge et al.'s theory on stress reaction of balance failure and balance recovery, also provides modeling and parameter computing scheme in cloud computing. On the basis, to complete a cloud computing resource schedule algorithm and simulation results approve the new algorithm is better than the Min-Min on balance load.

References

- [1] Baomin Xu, Chunyan Zhao, Enzhao Hu, Bin Hu. Job scheduling algorithm based on Berger model in cloud environment. *Advances in Engineering Software*, 2011, 42(7): p.419–425.
- [2] Y. Lu, Y. Chen, R. Latham and Y. Zhuang. Revealing Applications' Access Pattern in Collective I/O for Cache Management. In the Proc. of the ACM International Conference on Supercomputing (ICS'14), 2014.
- [3] Y. Lu, Y. Chen, R. Thakur, and Y. Zhuang. Memory-Conscious Collective I/O for Extreme Scale HPC Systems. In the Proceeding of the 3rd International Workshop on Runtime and Operating Systems for Supercomputers (ROSS), in conjunction with International Conference on Supercomputing (ICS), June 2013.
- [4] Y. Lu, Y. Chen, P. Amritkar, R. Thakur, and Y. Zhuang. A New Data Sieving Approach for High Performance I/O. In the Proc. of 7th International Conference on Future Information Technology (FutureTech'12), Vancouver, Canada, 2012.
- [5] H. Wu, M. Rahman, J. Wang, U. Louderaj, W. L. Hase, and Y. Zhuang, Higher-accuracy schemes for approximating the Hessian from electronic structure calculations in chemical dynamics simulations, *THE JOURNAL OF CHEMICAL PHYSICS*, 2010, 133, 074101.
- [6] Heng Wu, Jianxin Wang, and Yu Zhuang., A Parallel Domain Decomposition Algorithm for Solving the Equation of Nitric Oxide Diffusion in the Nervous System, (PDPTA'11).
- [7] Yu Zhuang, Heng Wu, Efficient Parabolic Solvers Scalable Across Multi-Architectural Levels (ISPA'12).
- [8] Heng Wu, Xuchong Liu, Ruihua Sun and Ningjia Zhu., ON THE CONVERGENCE OF A NONOVERLAPPING DOMAIN DECOMPOSITION METHOD WITH ROBIN-TYPE INTERIOR BOUNDARY, *INFORMATION, An International Interdisciplinary Journal*, 2014, Vol.17, No.8,

p.3547-3558.

[9]Wu heng, Jianxin Wang,Xuchong Liu,Eduardo Colmenares,Guofeng Yan,The Time Accuracy Analysis of Crank-Nicolson Predictor-corrector Numerical Scheme for Diffusion Equations, International Journal on Numerical and Analytical Methods in Engineering (IRENA),vol.1 no.2 April 2013,p.123-126.

[10]Zhao Chunyan, The Research and Realization of Job Scheduling algorithm under Cloud Computing, Beijing Traffic University, Master thesis,2009.

[11]Berger, J.,B.P.Cohen, T.L.Conner, M.Zelditch. Status Characteristics and Expectation States: A Process Model*[M]. Sociological Theories in Progress(Volume.One). Boston: Houghton-Mifflin. 1966, p.47-74.

[12]Berger, J., M. Zelditch, Jr., B. Anderson, and B. P. Cohen. Structural Aspects of Distributive Justice: A Status Value Formulation[M]. Sociological Theories in Progress(Volume Two). Boston: Houghton-Mifflin. 1972, p. 119-146.

[13]He, XiaoShan, XianHe Sun, and Gregor Von Laszewski. QoS guided min-min heuristic for grid task scheduling.Journal of Computer Science and Technology 18, no. 4 (2003): p. 442-451.

[14]Calheiros, Rodrigo N., Rajiv Ranjan, Anton Beloglazov, César AF De Rose, and Rajkumar Buyya. CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms.Software: Practice and Experience, 2011, 41(1): p.23-50.