

Diversification of Cloud Resource Allocation based on Improved Genetic Algorithm

Jian Wang¹, Wenyang Ke¹, Ying Yin^{*}, Bin Zhang

College of Computer Science and Engineering, Northeastern University, Shenyang, China

¹These authors contributed equally to this work and share first authorship

Corresponding author: yinying@mail.neu.edu.cn ;

Keywords: cloud resource allocation; Service-based Software system; genetic algorithm.

Abstract: Considering the optimization and multiple service levels agreement, this paper converts the resource allocation problem of Service-Based Software system in the cloud environment into a multi-objective optimization problem. Aiming at the cloud service resource allocation problem, this paper establishes a cloud resource optimization allocation model, and proposes a diversity elastic resource allocation algorithm based on genetic algorithm to solve the problem. To improve the efficiency, Algorithm uses integer coding, and in the selection operator, the elite reservation strategy and tournament strategy are introduced to ensure the convergence to the global optimal solution. In order to improve the global searching ability of genetic algorithm and to speed up the convergence rate, this paper adopts the adaptive mutation operator. Experiments verify the proposed resource optimization assignment model and solution algorithm effective, and show that the improved genetic algorithm can rapidly acquire with lower cost and higher resource utilization efficiency of the resources allocation strategy in large scale.

1. Introduction

Cloud computing environment provides users with unlimited computer resources and grants them to obtain and release cloud resource in a short time[1,2]. However, the dynamic resources and flexible need make the predesigned strategies of allocating resources deteriorate during the change of users' need and environment. Besides, in cloud computing environment, the existing pricing and scheduling strategies neither absolve conflict of interest between users nor give any cost-performance guarantee; meanwhile, there is a conflict between users' demand of lowest resource cost and shortest time of task completion and cloud service providers' expectation of maximized resource utilization rate and profit. But the undesirable conflict of interest can be well solved through resource allocation strategies. Thus, optimization of resource allocation strategy is one of the biggest concerns of cloud service provider[3].

The previous works on cloud resource allocation main come from the perspective of workflow and business process. Zhu [4] uses cloud services to realize scientific workflow, and proposes a resource scheduling method based on particle swarm optimization, and to minimize the overall cost as the goal. Pandey [5] uses performance objectives into component level goals, and to determine the resource requirements of each component, and the validity of the method is proved using multi-tier application. Schulte [6] put forward ViePEP elastic process platform, combining the function of business process management system with the cloud resource management system, capable of scheduling the process or a single task to meet the service level objectives. At the same time as much as possible to reduce the rate of resource cost and idle.

Different from those works, this paper is under the state of cloud resources solution to meet the SLA constraints and resource cost optimization target of the SBS resource allocation strategies. However, previous works focus on QoS optimization without considering SLA constraints firstly. Secondly, this paper uses the service selection strategy as a basis for solving of SBS in cloud environment resources optimization allocation problem.

The major contributions of this paper include: (1) proposed an energy-saving cloud resource allocation model based on SBSE; (2) designed a diversified and flexible resource allocation based on improved genetic algorithm; (3) evaluated amount of experiments with proposed strategy.

2. Solution to SBS cloud resource allocation

Suppose there is a SBS consisting of m component services and a cloud environment containing n physical machines, as shown in Figure 1, where the allocable resource amount of each physical machine is $R_k^\alpha (1 \leq \alpha \leq 2, 1 \leq k \leq n)$. Obtain the resource allocation set of each component service $R_k^\alpha (1 \leq \alpha \leq 2, 1 \leq k \leq n)$, where $C_i = \langle r_i, c_i \rangle$, resource state vector $r_i = (r_i^1, r_i^2)$ and resource cost $c_i = \sum_{\alpha=1}^2 c_i^\alpha$. As suggested by the resource pricing model, cloud resource is charged by the use time; therefore, an optimal allocation of SBS cloud resource on an energy-saving basis refers to the selection of sets of alternative resource allocation suited to component services, to maximize the resource utilization rate, minimize the total resource cost and satisfy the constraint of end-to-end response time.

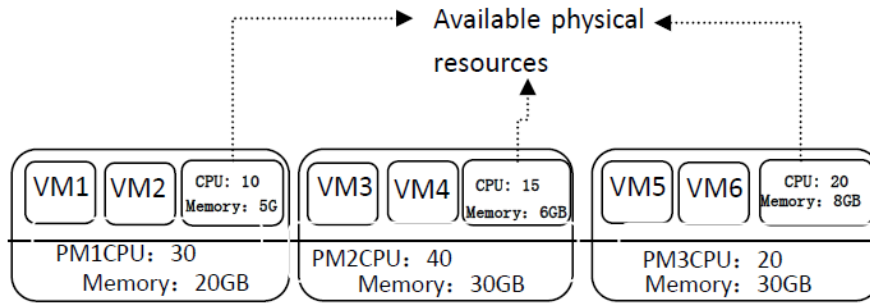


Fig.1 state of available resources of the physical machine

According to the above definition for the issue, the formal definition of cloud resource optimization allocation model is as follows:

$$\text{Min Cost} = \sum_{i=1}^m c_i (1)$$

$$\text{Max } U = \sum_{j=1}^n U_j (2)$$

$$\text{s.t.} \begin{cases} \text{Load}(S_i) < L_i (3) \\ \sum \text{Cost}(S_i) (4) \\ r_1 \leq C_c (5) \\ r_2 \leq M_f (6) \\ R(S_i) \neq R(S_j) \& T(S_i) \neq T(S_j) (7) \end{cases}$$

In the formula (1), Cost denotes the total cost of allocating resource for every service component in SBS; Formula (2) describes the occupancy rate of resource having been allocated to available resources before allocation; Formula (3) represents that load of component s_i cannot exceed L_i ; Formula (4) represents that the cost of component resource cannot exceed C; (5) and (6) represent, respectively, maximum available CPU and RAM resources allocated to each physical machine. (7) means the resources cannot conflict in time and space.

In practical application, the pricing model is used to the cloud calculate the cost of resources for service provider business. For convenience of description, this paper uses linear pricing mode for all types of resources. That is to say, it is assumed that the resource cost is proportional to the number of resources and duration of resources use. Then, the cost of resources $\alpha (1 \leq \alpha \leq 2)$ is:

$$c^\alpha = c_{price}^\alpha \times c_{amount}^\alpha \times t_{duration} (8)$$

In formula (8), c^α denotes the cost of resources α within components (α can be the CPU and RAM) and c_{price}^α denotes resource α 's price within the unit time. c_{amount}^α denotes the number of resources α used within components. $t_{duration}$ denotes duration for component taking up resource α .

The resources on the physical host are divided into two states: use state and idle state. The resource utilization rate of U in this paper is:

$$U_f = \frac{U_f^1 + U_f^2}{2} \quad (9)$$

In formula (9), U_f^1 and U_f^2 respectively represents utilization ratio of CPU and RAM of machines, where $U_f^\alpha = \frac{\sum c_{amount}^\alpha}{c_{total}^\alpha}$.

Definition (feasible solution to SBS cloud resource allocation): to an energy-saving and low-cost end by reducing the utilization rate of cloud resource. On the basis of optimal cloud resource allocation model, this paper uses modified genetic algorithm to reach an energy-saving resource allocation strategy up to SLA constraints with minimum resource cost, i.e., to specify in resource pool R a resource set that meets task T to reach maximum resource utilization rate and minimum total cost.

3. Resource allocation based on modified genetic algorithm

Based on software engineering thoughts and consideration of multiple grades of service, this paper converts SBS cloud resource allocation into an issue of multi-objective optimization and tries to solve it with meta-heuristic search algorithm. Genetic algorithm is a computing model that mimics Darwin's biological evolutionism and biological evolution of genetic mechanism, as well as a meta-heuristic search algorithm through which the optimal resolution is reached based on the simulation of natural evolutionary process. To provide a better solution, this paper takes advantages of particle swarm optimization and modifies the existing genetic algorithm to determine appropriate evolution strategies and fitness functions. As thus, a new genetic algorithm is developed to realize the global search of the solution space.

Fitness function evaluates individuals. According to the optimized cloud allocated model proposed above, in addition to having individuals meet the end-to-end response time and resource constraints of the physical machine, experimenters should take a full consideration to the cost of the allocation scheme. Since there are certain rules between components, this paper uses flexibility to evaluate the stability (i.e., the adaptability to emergencies) of the resource allocation scheme. This paper sets threshold values for partial variables to reduce the fitness of individuals without corresponding feasible solution in the solution space as so to prevent them from being transmitted to the next population and accelerate the algorithmic convergence. Fitness function is:

$$Fitness(X) = \theta_1 \frac{\sum_{i=1}^m c_i}{C} + \theta_2 \sum_{f=1}^n U_f + \theta_3 E(X) \quad (10)$$

In the formula (10), $\theta_1, \theta_2, \theta_3$ respectively is the coefficient of cost, resource utilization and the degree of elasticity. To ensure that the allocation of resources does not conflict, the cost of the threshold is set for C, and U_f is always less than 1. Individuals beyond the threshold will be discarded. E (X) is the degree of elasticity of individual X, which is determined by the average free time of parallel components.

4. Experiments Analysis

This paper conducts Matlab2014a simulative experiment to testify the validity and high efficiency of optimized SBS resource allocation model and modified GA in searching energy-saving scheme with high resource utilization rate and low cost. The experiment includes (1) introduction to experimental environment;

In order to describe the optimal resource allocation model, we only consider CPU and memory in the Qos attribute to two factors, setting the number of physical hosts as 5 and the number of components as 10. The physical machine resources are shown in Table 1:

Table 1 physical machines available resources

Host Resource type	A	B	C	D	E
CPU	20	22	21	22	20
Memory (G)	18	19	17	21	19

The group of experiments take the number of SBS system components as an example, and research the efficiency in solving the abstract component service model for the resource allocation strategy, evaluating minimum resource cost and physical machine resource utilization. The experimental results are shown in the Table 2:

Table 2 Effects of different service components

Number of service components	HGA execution time (s)	Minimum resource cost	Physical machine resource utilization
6	1.123	886	0.486
7	1.651	1070	0.559
8	2.879	1045	0.531
9	12.680	1340	0.537

According to the table, in the case of the same physical host, the algorithm has better performance in processing the 8 components, and the obtained solutions are better in resource cost and resource utilization. Analyzing the phenomenon when the component number of nine or more the algorithm execution time is too large, we come to the cause that the greater the number of components, physical machine resources competition is more intense, prone to physical resource conflict, causing individual solution abandoned, need to repeat the calculation. The physical machine state is given in this paper, which is suitable for 6-8 components. In the future, we think it can be used in the practical application.

5. Conclusion

This paper discusses a kind of resource optimization allocation method for SBS in cloud environment. The method is based on the SBSE idea, which converts the resource optimization allocation strategy to the service selection problem, and uses the improved genetic algorithm to search the optimal solution. Experiments show that the preference of the proposed method

ACKNOWLEDGMENT

Project supported by the National Nature Science Foundation of China (No. 61272182, 61572117), State Key Program of National Natural Science of China (61332014) and Fundamental Research Funds for the Central Universities (N150402002, N150404008).

References

- [1] Harman M, Mansouri SA, Zhang Y. Search-based software engineering: Trends, techniques and applications. *ACM Computing Surveys*, 2012, 45(1):11-61.
- [2] Zeng L, Benatallah B, Ngu AHH, Dumas M, Kalagnanam J, Chang H. QoS-aware middleware for Web services composition. *IEEE Trans. on Software Engineering*, 2004, 30(5):311-327.
- [3] Rao J, Wei Y, Gong J, Xu CZ. QoS guarantees and service differentiation for dynamic cloud applications. *IEEE Trans. on Network and Service Management*, 2013, 10(1):43-55.
- [4] Zhu Q, Agrawal G. Resource provisioning with budget constraints for adaptive applications in cloud environments. *IEEE Trans. on Services Computing*, 2012, 5(4):497-511.
- [5] Pandey S, Wu L, Guru SM, Buyya R. A particle swarm optimization-based heuristic for scheduling workflow applications in cloud computing environments. In: *Proc. of the 24th IEEE Int'l Conf. on Advanced Information Networking and Applications (AINA)*. Perth, 2010. 400-407.

- [6] Schulte S, Schuller D, Hoenisch P, Lampe U, Steinmetz R, Dustdar S. Cost-driven optimization of cloud resource allocation for elastic processes. *International Journal of Cloud Computing*, 2013, 1(2):1-14.