

Ant Colony Optimization Computing Resource Allocation Algorithm Based on Cloud Computing Environment

Guo Xin

School of Software Service Outsourcing
Jishou University
Zhangjiajie, China
jianghai079@126.com

Abstract—Ant colony algorithm is a kind of intelligent bionic algorithm. It is based on the observation of ant's food hunting in the nature, which finds that ants always find shortest path between their nest and food source. The main principle is ants can release pheromone and percept its concentration to identify and determine the next step, and release a certain concentration of pheromone according to the path length after food source is founded. The higher the pheromone concentration of the path, the greater chances of the path is choose. With positive feedback mechanism and robustness, ant colony algorithm is widely studied and improved since put forward. As ant colony algorithm shows great advantage in solving combinatorial optimization problems. Through the simulation analysis and comparison of grid, this algorithm can meet the requirements of the cloud computing environment, and get more response time and better quality, which is more suitable for cloud environment.

Keywords- cloud computing; grid; ant colony; resource allocation; algorithm analysis

I. INTRODUCTION

Cloud computing (Computing Cloud), is a super computing model which is connected through the Internet, including Computing Distributed, Computing Parallel (Grid), and the grid computing (Computing) technology, or the business implementation of these concepts of computer science^[1-2].

A new type of shared infrastructure, cloud computing, can be connected to a huge pool of systems, in order to provide users with a variety of storage and computing resources through the Internet. In the cloud computing environment, the user will own a large amount of information and processor resources on their personal computers, PDA or mobile phones and other devices to focus on, work together[3]. This is a large-scale distributed computing model, which is determined by the economic scale of the operators, and is abstract, virtual and dynamic variable[4]. Its main content is the ability to manage the computing power, storage, platform and services. These elements are distributed to the external users through the Internet, and the important meaning is to make the computing power as a commodity in the Internet[5].

The main advantages of cloud computing: the rapid reduction of hardware costs and enhance the computing power and storage capacity, the user can be very low cost

to get a high computing power, and no longer invest to buy expensive hardware equipment, the burden of frequent maintenance and upgrade[6].

Computing resource allocation[7] is an important part of cloud computing technology, which directly affects the performance of the entire cloud computing environment. Because of the unique characteristics of cloud computing, the resource allocation and scheduling algorithms for grid computing have been unable to work effectively in this environment[8]. The ant colony optimization allocation algorithm is proposed in this paper, which takes into account a series of characteristics of cloud computing, in order to efficiently allocate the appropriate computing resources for users in this environment[9].

II. PROBLEM DESCRIPTION

Cloud computing has evolved from grid computing and the grid computing as its backbone and basic structure. It can be said that cloud computing is a more advanced form of grid computing. However, there is a huge difference between the two, which can be found in the literature.

Cloud computing provides more abstract resources and services. These resources and services can be divided into three categories, namely a (as Platform Service Software), as a Service [2 (Service a) and 3] (as), Infrastructure.

In software, the service (SaaS), the user will receive a special purpose client, which allows the user to remote access through the Internet, and based on the use of the cost.

In the platform, service (PaaS), operators provide a high level of integrated environment to generate, test, and deploy the user defined application. This integrated environment is a kind of encapsulation of the development environment and the effective service load.

In the device (IaaS), the service quality agreement (Agreement Service-Level), which is achieved by the operator and the user, by providing hardware, software and equipment (mainly in the unified resource layer, but also in the organization layer), to provide a complete software should use the environment, and to charge the use of resources. In the level of service, according to the elastic characteristics of the cloud computing environment, the size of the device will dynamically increase or decrease according to the user's demand for resources. For example, users need to use T process, cloud computing environment

is allocated to the user R computing resources, such as the number of users will increase to 2T, then the user's computing resources will be extended to 2R. This paper discusses the calculation of resource allocation strategy, and also takes into account the characteristics of cloud computing[10].

At the level of the device, the user's all data and configuration are encapsulated as a user image, which is stored in the cloud environment. Due to the operational constraints, the network available bandwidth of the cloud computing environment is much smaller than the grid environment. For example, a traditional enterprise computing facilities generally have a super fast Ethernet 10G, high performance computing resources will be directly connected to the storage area network (Area Network Storage) through the network. And in the cloud computing environment, such as Amazon's EC2 (Compute Cloud Elastic) system only provides the available bandwidth of 250MB, and all the virtual servers are accessed through a single 250MB S3 Simple Service (Storage). Therefore, in the cloud computing environment, the local bandwidth needs to be fully utilized, a effective method is to try to allocate the local resources for the storage node resources. For this paper, the algorithm is to try to provide the user with the image of the local storage node to allocate the local or adjacent to the demand for less bandwidth computing nodes.

One of the operational objectives of the cloud computing environment is to serve as a service to the user to buy. Enterprises and individual users need not to buy expensive hardware acquisition costs, only through the Internet to buy lease computing power, the personal computer as the access port, everything connected to the cloud computing environment through the Internet, to complete the acquisition of computing power. However, this also indicates that the scale of cloud computing will be very large. At the same time, it is different from the grid's exclusive resource allocation model, and the resource in the entire cloud domain will be shared by all the users, in order to guarantee the delay sensitive operation in the cloud can be a good run. This means that the user's work in cloud computing will be divided into processes and even thread granularity levels. So, cloud computing has a strict requirement for computing resource allocation algorithm. This paper focuses on the calculation of resource allocation algorithm, which is to maximize the use of cloud computing environment to improve the effective utilization of computing resources through the most appropriate computing resources.

III. ALGORITHM DESCRIPTION

A. Computational resource allocation process

Referring to the cloud computing framework proposed by [6-7], Map/Reduce, each cell in the cloud environment is composed of a single Task Tracker master (Job Tracker) and one of the nodes in each cluster. The master node is responsible for scheduling the tasks that constitute a job, and the data resources are distributed among different users from the node's memory resources, the main nodes monitor their execution, and the task of re execution has

failed. And from the node is only responsible for the implementation of the tasks assigned by the master node. After receiving the task assigned by the master node, the node starts from the node to find a suitable computing node for the storage nodes. First, the node starts to detect its own computing resources, if the remaining computing resources are sufficient to meet the user's amount of work, the priority allocation of its own computing resources, if the resource has been exhausted or insufficient to meet the minimum amount of computing resources promised to the user, then began to search for the appropriate other computing resources in the cloud environment. The ant colony algorithm is introduced in this paper. Search is carried out in a certain range, and the purpose is to reduce the network overhead. If there is still no adequate resources, from node submitted to the main scheduling nodes move away the node in the cluster user data mirror slice.

B. Computing resources and the evaluation of the merits of the condition

The slave node field is considered as a $G(V, E)$, where V is the collection of all slave nodes in the region E , Area is a collection of slave nodes. Appropriate computing nodes, looking for is in the E , finding an optimal path E in area, the metric can consider the following factors:

Expected execution time: $\text{time_cost}(E)$, computing resources at the end of the road to the end of the e to deal with the waste time of the job is expected.

Network bandwidth: $\text{bandwidth}(E)$, the maximum bandwidth provided by the E .

Network delay: $\text{delay}(E)$, the maximum network delay generated by E . Set the constraint function for resource selection:

$$\text{res}(e) = \frac{A\text{time_cost}(e) + C\text{delay}(e)}{B\text{bandwidth}(e)} \quad [1]$$

$$s.t. \begin{cases} \text{time_cost}(e) < TL \\ \text{bandwidth}(e) > EL \\ \text{delay}(e) < DL \end{cases} \quad [2]$$

IV. ALGORITHM SIMULATION

As a part of the cloud environment can be viewed as a special grid, we use Gridsim to simulate the local domain in a cloud, to check the operation of the algorithm in this particular grid environment. Among them, we use the annealing algorithm and genetic algorithm as a comparison. Through the GridResource class and a series of auxiliary classes in Gridsim, we simulate the computation and network resources of cloud computing, in order to build a more realistic cloud environment. To manage resources by setting the GridInformationServices class. By using two to determine the quality of resources, we can compare the performance of these algorithms by constructing different scheduling methods in GridBroker.

V. REDUCE NETWORK LOAD

When a forward ant enters a slave node, it sets itself into the avoidance node set of the ant, and calculates its

own effective nodes by limiting the conditional equation and formula (3). If the slave node satisfies the limiting condition, the tag is labeled as an effective node, modifies the corresponding pheromone intensity, and generates a backward ant. Carry forward the information of the line and the computing resource to the ant colony to return, and the last node of the path node set is to find the effective joint. If the slave node is not an effective node, the forward ant continues to step 2. In order to prevent the algorithm from rapid convergence, a node can set a frequency threshold F , if a path before to the ant access frequency exceeds the threshold, with a certain probability p to produce after to the ant.

(4) the nodes will be amended to limit the conditions of this, the network factor will be bandwidth minus the amount of pre allocated occupancy. If the rest of the network is not sufficient to meet the operating conditions, then let in the node within the end, otherwise let go and search.

Each pheromone is not zero, according to the formula (6), (7) by the time interval to update their own pheromone intensity.

(6) at the present time () when the $H_p <$ value of the ant's life value is Hops or is unable to continue to go (such as when a neighbor node is not connected to the second node), the ant automatically dies. Among them, the value of the ant's H_p actually control the search area.

(7) to the nodes in the transmission bandwidth to collect along the way. The source node is produced by the information of the ant colony, and the effective resource set is ordered.

(8) when the computing resource is found to be insufficient to meet the needs of the user, the first step is to be transferred to the first step. Otherwise, the algorithm, according to the effective resource set will be assigned to the job to calculate the resource node, and as high as possible to the high availability of nodes to allocate more work, to reduce the load of the network.

(9) when the algorithm stops, all computing resources are not sufficient to meet the user requirements of the agreed service level agreement (SLA), then consider the transfer of the user image of the storage node to other storage nodes.

A. Figures and Tables

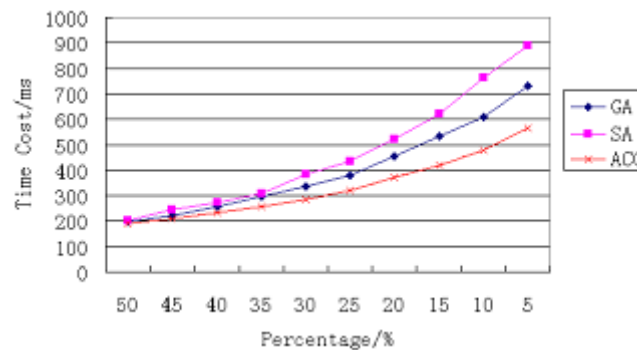


Figure 1. Graph of simulation result

Positioning Figures and Tables: In Figure three, the ratio of the horizontal coordinates to the effective nodes, the longitudinal coordinates of the 5 algorithms is the time consuming of the successful allocation of the active nodes. From the results, the algorithm is more effective and less efficient in the case of the other two algorithms, which is the characteristic of the cloud environment. So this algorithm has more advantages in cloud computing environment.

Table 1. Main parameters of these algorithms(2)

Algorithm	Project	Value
GA	Population size (α)	25
	Cross ratio	0.5
	Variation rate	0.03
SA	Cooling coefficient	0.25
	Step by step before cooling (step)	
	Cooling step size (degree)	
	Starting temperature)	

ACO	Ant colony size α	25
	β	2
	ρ	6
		0.15

VI. CONCLUSIONS

Due to the operational requirements of the cost, cloud computing is less than the grid environment, and the service content such as Saas, Paas and other services is higher than the grid, so a good resource scheduling algorithm is very important. This paper introduces the ant colony resource allocation algorithm for large-scale, sharing and dynamic characteristics of the cloud environment, dynamically for the user's job search and allocate computing resources. From the simulation results, it can be seen that this algorithm can effectively accomplish the work of searching and allocating resources in the cloud computing environment.

ACKNOWLEDGMENT

A Project Supported by Scientific Research Fund of Hunan Provincial Education Department.

Project number: 14B143.

REFERENCES

- [1] Giacomo V. Mc Evoy and Bruno Schulze. Using Clouds to address Grid Limitations[C]//MGC'08. Belgium:Leuven Press. Dec 1-5,2008.
- [2] Ian Foster, Yong Zhao, Ioan Raicu and Shiyong Lu: Cloud Computing and Grid Computing 360-Degree Compared[C]//Grid Computing Environments Workshop.IEEE.12-16 Nov. 2008.
- [3] Huan Liu and Dan Orban, Accenture Technology Labs .GridBatch: Cloud Computing for Large-Scale Data-Intensive Batch[C]//CCGRID,2008.30 8th IEEE International Symposium on (2008). Shanghai:[s.n.], 2008.
- [4] Amazon Elastic Compute Cloud (Amazon EC2), <http://aws.amazon.com/ec2>, 2008.
- [5] Amazon Simple Storage Service (Amazon S3). 2008 Sep. <http://aws.amazon.com/s3>.
- [6] Chen Kang, Zheng Weimin, Cloud Computing: system examples and research status [J], software journal, 2009, 20 (5): 1337-1348.
- [7] Zhang D, Guo D, Chen F, et al. TL-plane-based multi-core energy-efficient real-time scheduling algorithm for sporadic tasks[J]. ACM Trans. Archit. Code Optim, 2012,8(4):47-48.
- [8] Yong Xian Li, Xu Lun Luan, Sen Sen Li. Cloud computing technology in the library application [J], the Journal of the Library Science in Jiangxi, 2009 (1): 105-106.
- [9] Zheng X, Cai Y. Energy-aware load dispatching in geographically located ineernet data centers[J]. Sustainable Computing: Informatics and Systems, 2011, 1(4):275-285.
- [10] Lu Xiaojuan, cloud computing and future library digital information resources construction [J], Sichuan Library Journal, 2009 (2): 23-24