Multi Objective Optimization of Virtual Machine Migration Placement Based on Cloud Computing

Sun Hong[1,2], Tang Qing[1],Xu Liping[1] and Chen Shiping[1]

[1]University of Shanghai for Science and Technology, Shanghai China 200093,

[2]Shanghai Key Laboratory of modern optical systems, Shanghai China 200093

Corresponding Author: Sun Hong, University of Shanghai for Science and Technology, 200093, shanghai, China Email:sunhong@usst.edu.cn

Abstract. How to improve the resource utilization of the cloud computing system has been one of the key content of the research of the cloud computing. The traditional multi-objective ant colony optimization was improved, studied the virtual machine live migration framework, combined with the elimination method to solve virtual machine migration and placement of multi-objective optimization problem, the load balanced specific strategies are integrated into the framework of a dynamic migration, simulation experiments are carried out and the conclusions are made for it. The algorithm can obtain the optimal solution through the continuous updating of pheromone. The main consideration is the Service level contract violation rate(S), Resource loss(W),Power consumption (P). Experimental results show that ,compared with the traditional heuristic method and genetic algorithm, the algorithm is advantageous to the parallel computation, and it’ s able to achieve the optimal tradeoff and compromise between multiple conflicting objectives. In the case of service level contract violation rate is low, system resource waste and power consumption are at the least, so it has feasibility.

Keywords: Cloud computing, Virtual machine migration, Multi objective optimization, Ant colony algorithm, Elimination method.

1. Introductiong

As a new technology, cloud computing has become a hot research topic in the field of information in recent years. As a new business computing model, business characteristics and virtualization technology is its obvious characteristics. And the task scheduling is the key technology of cloud computing, it not only affects the efficiency of the whole system, but also significantly affects the quality of service. At present, the problem of task scheduling is studied in the grid environment. Due to the diversity of user requirements in the cloud computing system, and the complexity of the task type, previous task scheduling algorithms can’ t meet the requirements of the overall QoS. In the dynamic cloud computing environment, improving the efficiency of task scheduling and load balancing is an eternal problem, for users, to meet the user’ s QoS requirements is the most important thing. Therefore, the research on the task scheduling algorithm with QoS expectation constraint is the key content of the

cloud computing system. Nowadays, cloud computing has become a new model to provide access and services through the Internet. If the distribution of resources is not reasonable, it will inevitably lead to waste of resources. It is of great significance to realize the multi objective optimization of virtual machine migration and placement in the present stage. Most researchers use the traditional heuristic method or genetic algorithm and other algorithms to make the virtual machine placement before; Although these algorithms can solve the problem of virtual machine migration in a certain extent, but these algorithms have their own limitations. For example, heuristic method can solve the problem of local optimal solution in virtual machine migration, but the method is short of the ability of global optimization. Although genetic algorithm has certain advantages in multi objective optimization, it can't make full use of the feedback information, so that the search is blind, and the efficiency of solving the optimal solution to a certain extent is relatively low. ant colony optimization, which also called ant algorithm, is a kind of probabilistic algorithm used to find the optimal path in the graph. Ant colony optimization is a simulated evolutionary algorithm, a preliminary study shows that the algorithm has many excellent performance. Compared with the results of genetic algorithm design, numerical simulation results show that ant colony algorithm has a new simulation evolutionary optimization method and its effectiveness and application value.

This paper introduces the management framework of the two layers of local management and global management, through this management framework is conducive to the migration of virtual machine placement and resource allocation to make better decisions. The method used in this paper is to improve the traditional ant colony algorithm, and combine with the exclusion method. It is advantageous to the parallel computation, and the efficiency is higher. To be able to obtain the optimal tradeoff between the three conflicting objectives which are service level contract violation rate(S),resource loss(W) and power consumption(P),and in the case of service level contract violation rate is low, the waste of system resources and power consumption are the least.

2. Overall Framework for Virtual Machine Migration

2.1 Virtual machine migration

Xen virtual machine is using virtualization technology which is a quasi virtualization technology, and have good performances in all kinds of architecture, it has very good performance and system isolation. Up to now, Xen is definitely the most outstanding Linux system under the open source virtual machine, Xen is now no longer originally just supported by x86, and now has wild support and even Itanium and other hardware platforms are available to it. Version 4 was released in 2010, the client can support up to 64 virtual CPU. To use a virtual machine, you should start the operating system, but Microsoft platform VMware is the first to enable the physical machine system and then start the process, and this is not the standard. After Xen started, the first step is to run the virtual machine monitor, which is Xen Hypervisor (also known as the super management program in the Xen system), then run the host operating system (or local operating system),by minimizing the connection between the super manager and the native operating system, it reduces the risk of super management program itself and the virtual machine being destroyed and information leakage.

2.2 Overall framework and optimization of virtual machine migration

The basic migration structure is implemented by four modules, including migration monitoring, execution migration, suspension and rewake. As shown in Figure 1.

Monitor migration module: The primary function of the primary module is to determine the source of the migration, the start time of the migration, and the purpose of the migration. The working mode of listening and migrating module is determined by the purpose of migration. In order to ensure the load balance of each node, setting up the monitor signal in the virtual machine management program, according to the monitoring of the various nodes of the load operation to determine whether the need to migrate. Set a migration threshold, when monitor to arrive at this value, monitor migration module will send a migration signal to the source, indicates that the source machine will be migrated. Meanwhile, monitor the migration module to communicate with other nodes, look for the lower load nodes, and determine the specific location of the destination machine.

Run migration module: This module is the most important module of virtual machine migration, almost bear most of the migration work. After running the migration, this module collects the running information of the source machine, at the same time to freeze the module by sending the "frozen" signal to the source machine. This process is the key part of the migration process, directly affect the migration process downtime and migration of the total length.

Freezing module: This module is mainly responsible for how to solve the continuous service problem. It makes users feel not interruptions.

Target domain wake-up module: The function of the target domain wake-up module is to determine the time to wake up the destination machine, also ensure that the weaken target machine is consistent of the source machine, and how to ensure that the service of the target area is connected with the source area. After it shutdown, the running module will copy the remaining memory pages, then send the "weaken signal" to the weaken module which on the target machine.

Interrupt connection device is a direct consequence caused by shutdown, peripheral device cannot connect to virtual machine, this, of course, will cause the external service is not timely or appear all kinds of transmission errors. In order to improve the effectiveness of migration, reduce the time of shutdown, increase the application rate of load, we propose an optimized virtual machine dynamic migration framework, this framework is also based on Xen.

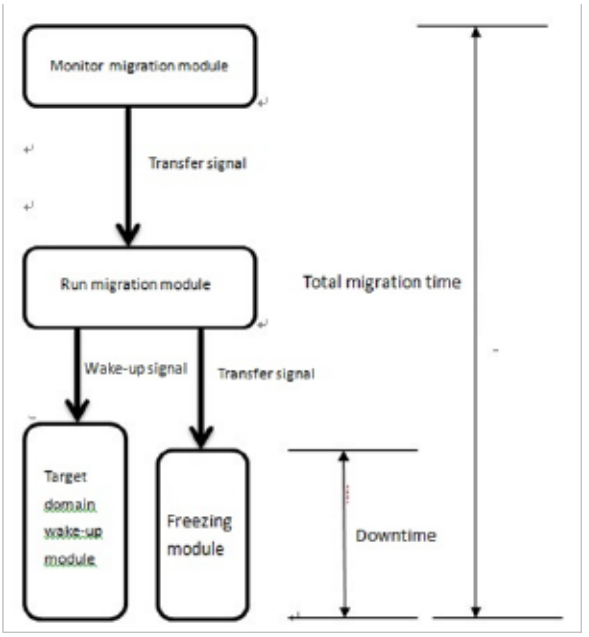

Figure.1 The virtual machine dynamic migration placement module

In order to improve the utilization of load, at the same time, it also makes the migration process more smooth and effective, optimization the design of dynamic migration of Xen virtual machine. Add two modules to achieve load balancing, one of them is the load monitor module, which add to the original monitor migration module, in order to set the identity of the current virtual machine running information, set the trigger conditions for its migration and prepare for the subsequent migration to select the appropriate load; the other one load transfer module is mainly responsible for the positioning and selection strategy of the virtual machine migration. As shown in Figure 2, three modules are marked by grey patterns.

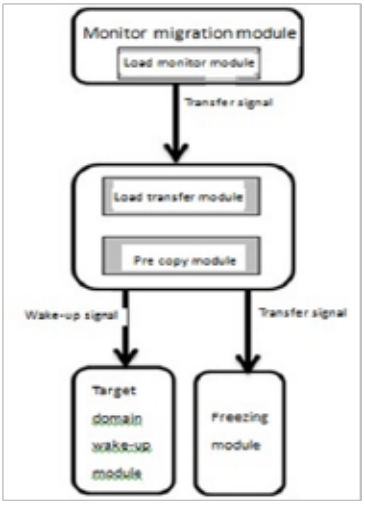

Figure.2 Dynamic migration of virtual machine placement framework optimization module

3. Virtual Machine Migration Placement Policy Model

A large number of computers will be integrated into resource pool in the cloud computing, then the computing tasks in cloud computing are distributed in the resource pool, all applications can each one takes what he needs, for example, you can have a stronger computing power to meet the needs of computing, you can set aside a larger storage space to store resources, and even more perfect online updates and other software services. In the cloud computing platform, a variety of servers to complete a specific task by way of collaboration. Because there is no unified standard for cloud computing platform application interface, the application of the various cloud environments can not be fully integrated, at the same time, the resources of the nodes in the cloud environment are different. For example, the formats provided by the same resource are different, the demands for resources in different time periods are different, there are large number of access to the same node during certain periods of time. This will lead a overload caused by excessive access on a server node in a certain period of time. While the other nodes are less load due to access relatively light, this forms the node utilization rate is not high in the whole system, causing the load is not balanced. Virtualization technology continues to mature, this is to provide a solution to this. The emergence of dynamic migration of virtual machine is an effective way to solve the load imbalance. The whole virtual machine running state can smooth and stable mutual transfer between the two physical hosts in the same cluster, of course, this is the necessary conditions for the transfers, and users don’t have any feeling of stagnation. The dynamic migration of virtual machine can assist the maintenance personnel of the cloud environment, so that the nodes in the cluster can be fully used, achieve load balancing dynamically. Therefore, to improve the resource allocation in the cloud environment and to strengthen the system by designing an efficient load balancing algorithm,

this has become one of the important issues in the field of cloud computing. Virtual machines allow all computing tasks to encapsulate into the virtual machine. Because the virtual machine is one of the characteristics of isolation, so you can use the virtual machine dynamic migration technology to migrate computing tasks. The scale of cloud computing is generally relatively large, it also provide the same size of the pressure to how to adjust the distribution of the node resources. Considering the real time information of resources in cloud environment, resource scheduling must be done, this requires real-time monitoring of resources in the cloud environment, and can dynamically manage. From the point of view of the process size of the task to be migrated, cloud computing users pay more attention to the migration process in the virtual machine itself how to operate, of course, the premise is to try not to affect the user. How to provide resources to the service level agreement of the internal application of the virtual machine is a problem to be solved at present. In terms of practical ability, resource scheduling system must monitor the usage of resources and provide reference for the system itself in time, or for system management related personnel to set.

3.1 Virtual machine management framework

The number of infrastructure nodes in cloud computing is very large, which makes it very difficult to build a structure. The management framework of this paper is two layers of management, local management and global management, the details are shown in Figure 3.The management of host cloud infrastructure to enable global management to run on a host node, by monitoring the collection of various information from local management, including user service quality resource consumption and power consumption, and so on, then make decisions on the placement of the virtual machine and the allocation of resources.

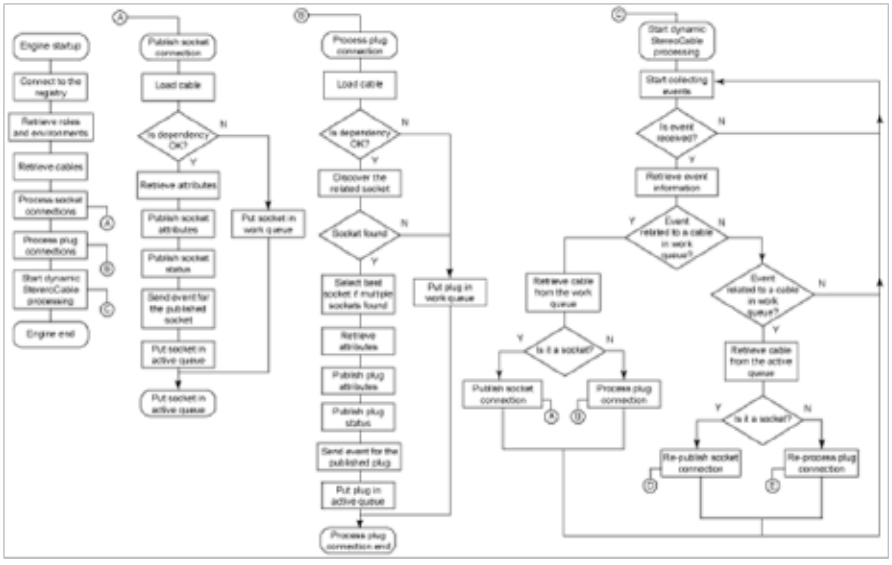

Figure.3 Virtual machine system management structure

3.2 Mathematical model of virtual machine migration

The migration and placement of virtual machine has been the focus of research in cloud computing, it is a typical bin packing problem. The literature proves that the virtual machine migration is a NP-hard problem. According to the framework of system management in Figure 3, in this paper, we need to consider the following 3 factors in the global management of virtual machine initialization: service level contract violation rate(S), resource loss(R), power consumption(P). m, n, respectively, indicating the

number of physical nodes and virtual machines, says j-th physical nodes have a corresponding resource capacity, represents the resource capacity of the request of i-th virtual machines. Its mathematical model is described as follows:

target:

$$\min \sum_{j=1}^{m} S_j \text{ and } \min \sum_{j=1}^{m} R_j \text{ and } \min \sum_{j=1}^{m} P_j \quad (1)$$

constraint:

$$\sum_{i=1}^{n} r_i^{CPU} \cdot a_{ij} < c_j^{CPU}, j = [1, \cdots, m] \quad (2)$$

$$\sum_{i=1}^{n} r_i^{mem} \cdot a_{ij} < c_j^{mem}, j = [1, \cdots, m] \quad (3)$$

$$\sum_{i=1}^{n} r_i^{bw} \cdot a_{ij} < c_j^{bw}, j = [1, \cdots, m] \quad (4)$$

$$\sum_{j=1}^{m} a_{ij} = 1, i = [1, \cdots, n] \quad (5)$$

The three objectives of the formula (1) are service level contract violation rate (S), resource loss (R), and power consumption (P). Formula (2-4) constrains the allocation of CPU, memory and network bandwidth resources on each physical node, which will not exceed the capacity of its own. And formula (5) constrains each virtual machine can only be assigned to a physical node.

4. Multi Objective Optimization Virtual Machine Migration Placement Strategy Based on Ant Colony Algorithm

Ant colony algorithm is a kind of technology that can be used to find the optimal solution. The algorithm is widely used in the virtual machine migration and placement problem, and has certain advantages in dealing with combinatorial optimization problems. The following is the specific design steps and process of this article.

4.1 Filness function

The selection of the stress function is very important in the genetic algorithm. According to the formula (1), the 3 sub - suitability function is defined, the value of the range is between [0~1] SLA violation rate function (f_{SLA}) ,resource utilization function ($f_{resource}$) ,power consumption function (f_{power}) , Such as formula (6) - (7).

$$f_{SLA}(u_{CPU}) = \frac{1}{1 + e^{u_{CPU} - 0.9}} \quad (6)$$

$$f_{resource}(u_{CPU}, u_{mem}, u_{bw}) = u_{CPU} \times u_{mem} \times u_{bw} \quad (7)$$

$$f_{power}(u_{CPU}) = \frac{u_{CPU}}{P_{idle} + (P_{busy} - P_{idle}) \times u_{CPU}} \times P_{busy} \quad (8)$$

In the formulas, u_{CPU} 、 u_{mem} 、 u_{bw} 、 P_{busy} indicate the CPU, memory, network bandwidth utilization and multiplier factor respectively on the physical node . f_{power} reflects the amount of effective work in a certain amount of power consumed.

Taking into account the need to balance the service level contract violation rate (S), resource loss (R), power consumption (P) 3 goals. So the weight value of this paper are set to 1, and according to the experience in this definition the suitability function is

$$f(u_{CPU}, u_{mem}, u_{bw}) = f_{SLA}(u_{CPU}) + f_{resource}(u_{CPU}, u_{mem}, u_{bw}) + f_{power}(u_{CPU}) \tag{9}$$

4.2 Pheromone

pheromone update rules as shown in formula (10) - (11)

$$\gamma_{iu} = (1-\rho) \times \gamma_{iu} + \Delta\gamma_{iu}^{best} \tag{10}$$

$$\Delta\gamma_{iu}^{best} = \begin{cases} f(S_{best}), JVM \text{ is loaded on the node u} \\ 0, \quad \text{others} \end{cases} \tag{11}$$

In the formula, S_{best} is the optimal solution set, ρ is the pheromone volatile coefficient, $\Delta\gamma_{iu}^{best}$ is the pheromone increment, $f(S_{best})$ is the appropriate degree function.

4.3 Probability transfer function

Probability transfer function

$$P_{iu}^{k}(t) = \begin{cases} \dfrac{\gamma_{iu}^{\alpha}(t) \times \eta_{iu}^{\beta}(t)}{\sum_{Seallowed} \gamma_{iu}^{\alpha}(t) \times \eta_{iu}^{\beta}(t)}, & i \in allowed_k \\ 0 \end{cases} \tag{12}$$

In the formula, is an information heuristic factor, is the visibility heuristic factor.

Among them, the gamma _i^CPU, gamma _i^mem, gamma _i^bw are virtual machine I request CPU, memory and network bandwidth of the corresponding resources on the host node u ratio.

4.4 The construction of the optimal solution set

Using the exclusion method to construct the non dominated set is a common method in multi-objective genetic algorithm. In this paper, we use the rule of law and its appropriate improvements to deal with the solution of the ant colony algorithm search process, which can be used to build the Paxeto solution set, the process is as follows:

Step1:Set the solution set $D_{cycle}^{*} = \{D_1, D_2, \cdots D_n\}$ for a loop search, where n is the number of the solution to the search.

Step2:To evaluate each solution vector has 3 sub goals, if D_i target is better than D_j corresponding to other sub goals and sub goals, D_i and D_j were compared to non inferior, concludes that D_i dominated D_j , D_j must be removed from the current set of solutions of C, and vice versa.

Step3： And so on, will the solution D_{cycle}^{*} were compared with each other, to get the optimal solution set D_{cycle}^{*} of cycles.

Step4: The D^{*}_{cycle} and the global optimal solution set D_{best} are compared according to the exclusion method, and the final non dominated solution is saved to the S_{best}.

Step5: To continue the cycle, when the cycle is over, the global optimal solution set D_{best} is the Pareto optimal solution set.

5. Experimental Results and Analysis

This experiment is done on the C1oudSim[6], cloudsim by Rajkumar professor Buyya team (Melbourne University) developed cloud computing simulator, Melbourne University in Australia Grid Laboratory and Gridbus project announced the launch of cloud computing simulation software.

CloudSim as a generic, scalable new simulation framework that supports seamless modeling and simulation, and can be carried out on the basis of cloud computing infrastructure and management services. This simulation framework has the following characteristics[7]:

Simulation and example of a large-scale cloud infrastructure supporting a single physical computing node.

To provide an independent platform, the main function is to the modeling of the data center, service agent and scheduling strategy.

In a data center node to provide a virtual engine, to manage the independent virtualization services.

Flexible virtualization services can switch between shared space and shared time processing core allocation policies.

In this paper, we use ant colony algorithm for resource allocation, and some other algorithms are compared.

Experiment set 80 physical nodes, each node is configured for GB l0 memory, 1TB storage and bandwidth of 1 Gbps, while the capacity of CPU is equivalent to 1000, 2000 and 3000 MIPS. The number of requests for a virtual machine is 200, where the request for CPU is 250500750 and 1000 MIPS, 4 GB memory, 200, GB bandwidth, 200 Mbps. The power consumption of the physical node in CPU utilization is 0% and 100%, respectively. The power consumption is 175W and 250W.

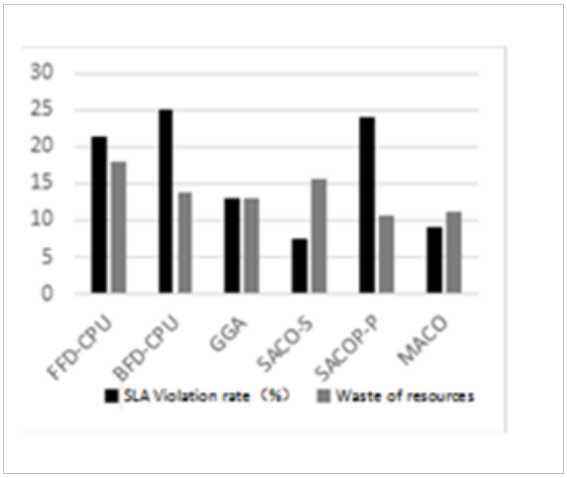

Figure.4 Comparison of SLA violation rate and resource waste in 6 placement algorithms

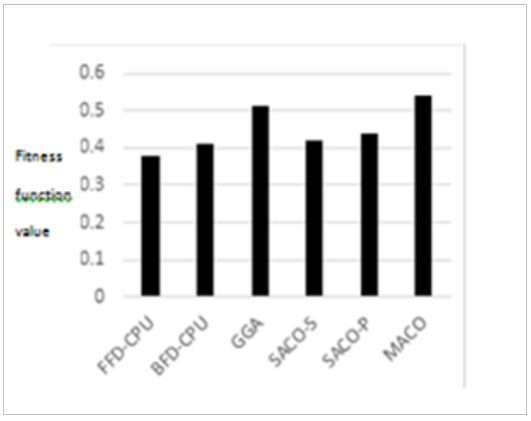

Figure.5 Comparison of the fitness function value of 6 kinds of placement algorithms

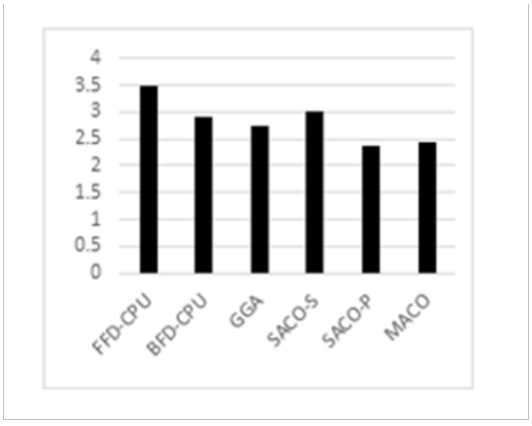

Figure.6 Comparison of 6 placement algorithms under power consumption

6. Concluding Remarks

The algorithm used in this paper is an improved ant colony algorithm for distributed multi objective optimization. This algorithm is an improvement of the traditional multi objective ant colony algorithm. Selected service level contract violation rate (S), resource consumption (W), power consumption (P) three targets. And combined with the elimination method to solve the virtual machine migration in the placement of these three objectives optimization problem. Experimental results show that compared with the traditional heuristic method and genetic algorithm, the proposed algorithm can effectively reduce the resource waste and the power consumption of the system when the service level contract violation rate is low, and it has feasibility. This paper has used the power consumption as one of the management objectives, next, we also need to consider how to take into account the data center network traffic and other aspects, so as to achieve more perfect.

References

[1] HYEAR C,MCKEE B, GARDNER R, et al. Autonomic virtual machine placement in the data center [J]. Hewlett Packard Laboratories, Tech. Rep. HPL-2007-189, 2007:2007-189.

[2] Li Jingchao,ChenJingyi,WuJie. Research on virtual machine placement based on improved grouping genetic algorithm [J]. Computer engineering and design,2012,33(5):2053-2056..

[3] Li Yong: Analysis and Research on dynamic migration of virtual machine[Dissertation]. National Defense

Science and Technology University,2007.

[4] Li Zhiwei,WuQingbo,TanYusong. Research on dynamic migration of virtual machine based on device agent mechanism. Computer application research. Twenty-sixth volumes, April 2009.

[5] CAREY M R， JOHNSON D S. Computers and ln tractability: a guide to NP-completeness [J].1979.

[6] CALHEIROS R N, RANJAN R, ct al. Cloud Sim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms [J]. Software: Practice and Experience, 2011, 41(1): 23-50.

[7] Sun Hong, Zhang Huaxuan, Chen Shiping, etal. The Study of Improved FP-Growth Algorithm in Map Reduce[C].//Proc.CPCI- 1st International Workshop on cloud Computing and Information Security(CCIS) 2013 CPCI:00033591040005

Acknowledgements

This work was supported by the National natural Science Foundation of China (No.61472256, No.61170277), Innovation Program of Shanghai Municipal Education Commission(No.12zz137), and the Hujiang Foundation(C14002).

Biographies

Sun Hong: female, Han, 1964-, from Beijing, China, Master, associate professor, School of Optical-Electrical and Computer Engineering University of Shanghai for Science and Technology, master tutor, associate professor direction of research; Business schools University of Shanghai for Science and Technology doctor graduate student; the main research direction: computer network communication and clouds computing, management science and engineering, Management Information and Decision Support System . Email:sunhong@usst.edu.cn,Telephone:13916902800

Tang Qing: male, han, 1993-, master student, School of Optical-Electrical and Computer Engineering University of Shanghai for Science and Technology; the main research direction: cloud computing and management information system.Email:qingtang1993@163.com

Xu Li-ping: female, han, 1986-, master, associate professor, University of Shanghai for Science and Technology; the main research direction: cloud computing and management information system. Email:5850487@qq.com

Chen Shi-Ping: male, han, 1964-,form Zhejiang, China, professor, Ph.D., doctoral tutor Business schools University of Shanghai for Science and Technology, research direction: computer network communication and clouds computing, management science and engineering. Email:chensp@usst.edu.cn