International Journal of Grid Distribution Computing
Vol.7, No.6 (2014), pp.201-210
http://dx.doi.org/10.14257/ijgdc.2014.7.6.16

Research on Optimal Scheduling of the Cloud Computing Resource based on the Genetic Algorithm in Distributed Computing Environment

Baoli Yuan, Bin Geng and Hongmei Sun

ShijiaZhuang Vocational Technology Institute, Shijiazhuang, Hebei, China
1705201@qq.com, gbfll@vip.qq.com, 335219117@qq.com

Abstract

In recent years, distributed computing technology has been one of the cutting edge technologies for its low power and cost, which makes numerous IT organizations extend their hands in order to improve their financial ability. Because of these new features, grid computing, the original task scheduling mechanism, can't work effectively in distributed computing environments, hence, we need a new task scheduling method to solve the problems. With considering the complex characters of the task in different distributed computing applications, firstly, we construct a more comprehensive task scheduling model, which has three sub objective functions. Secondly, we present an improved genetic algorithm to solve the multi-objective NP-hard problem. Finally, we implement some simulation experiments, and the evaluation results show us that the proposed model and improved GA are efficient and effective. The first part is the research status and related problems. The second part is the establishment of system architecture and task scheduling model. The last part is the experimental analysis and conclusion.

Keywords: *Distributed computing; task scheduling; task decomposition; genetic algorithm*

1. Introduction

Distributed computing is a technology that numerous IT organizations extend their hands in order to improve their financial ability [1]. This is achieved by improving the various parameters such as performance, throughput, reliability, scalability, load balancing, persistence and so on [2-3]. The services such as disk storage, virtual servers, application design, development, testing environment are added advantages of the distributed computing technology. As an important part of distributed computing, task scheduling is a mapping mechanism from users' tasks to the appropriate selection of resources and its execution. Report to grid computing, distributed computing has many unique features including virtualization and flexibility. By using the technology of virtualization, all physical resources are virtualized and transparent for users. All users have their own virtual device, these resources do not interact with each other and they are created based on users' requirements. In addition, one or more virtual machines can run on a single host computer so that the utilization rate of resources has been effectively improved [4]. The independence of users' application ensures the system's security of information and enhances the availability of service [5]. Supplying resources under the distributed computing environment is flexible, we increase or reduce the supplying of resources depends on users' demand. Owing to these new features, grid computing, the original task scheduling mechanism, can't work effectively in distributed computing environments [6]. Hence, we need a new task scheduling method to solve the problems.

ISSN: 2005-4262 IJGDC

In recent years, network, information, management, and other technologies and theories have been fast developed and widely applied. At the same time, the growth of competitive market globalization and customer demand diversification has led to the increasing demand of agility, networking, service, green, and socialization of computing. Lots of research works have been conducted on various aspects of distributed computing including protocol and architecture, network topology, routing and application prospects, power conservation, *etc.* [7, 8-10]. The operation of a distributed computing system can be divided into five stages: user request, resource exploration, task scheduling, service and process monitoring and returning feedback, in which the third one is the most important part because it directly influences the final quality of services and total cost during the process. Task scheduling supported in distributed computing environment is still remained as an open field of research from diverse perspectives. At present, there is no longer a unified standard and research framework, according to the different objective and context of task requirements. The literature presented a chief according to demand of task decomposition, and gave the task allocation and coordination technology for virtual task [11]; with the least total cost as the objective function, mm established a task allocation model on the processor in a distributed computing system and designed the improved genetic algorithm [12]; Guo and Zhao considered that same tasks can be divided and other tasks cannot be divided, two kinds of multi-processor optimal task assignment problem was studied. Taking the maximum and minimum cost as the objective function, the author put forward the approximate solution [13]. Liu H, Abraham A established the multi objective decision model and algorithm for selecting partners of virtual task [14], The literature on research of solution has a certain reference value, but only considers a single target [15-16]; Dutta D. *et al.*, proposed a market-based mechanism to allocate resources in a cloud computing environment, where the resources are virtualized and delivered to users as services [17]. For computing resource allocation, traditional researches mostly focus on the modeling and evaluation of computing resources based on homogeneous/heterogeneous cluster systems or distributed grid computing systems. User's demand for resources, the cost and computation and communication capabilities of resources are the major considerations among these studies. In distributed cloud computing mode, virtualization is the main support of flexible resource sharing. In this context, Gorbenko et al. introduced the concept, classification of resource allocation in distributed cloud [18]. Zhao *et al.* investigated the management of cloud computing resources based on ontology and virtualization respectively [19]. Konugurthi P *et al.* proposed a method for the deployment of upper layer software computing services from virtual machines [20]. Zhong H *et al.*, have studied the feedback-based optimization problem including the allocation of resources especially in private cloud [21]. Most of these studies concentrated on the expansion of characteristics of computing resources based on traditional models and the algorithm designing for task dispatch in distributed computing environment.

In summary, current research of task scheduling distributed computing environments is in its infancy [22], there is considerable problem space to explore and solve. However, the mutual relations between computing services and the underlying computing resources and the influence of virtualization on quality of services, as two of the key factors in distributed computing system, have not been studied. Meanwhile, previous studies often consider only one objective function, disregarding the complexity and diversity of user' computing task. Therefore, with considering the complex characters of task in different distributed computing applications, firstly, we construct collaborative computing system architecture, which includes task plan, global scheduling, collaboration allocation, virtualization and computing resource. Secondly, we propose a more comprehensive task scheduling model, establish a fuzzy analytical hierarchy model for the collaborative computing task assignment problem to satisfy different users' request, and then the weights of each objective are obtained. Finally, we implement some simulation experiments, and the evaluation results show us that the proposed model and improved GA are efficient and effective.

2. Collaborative Computing Task Scheduling Model

A In this paper, the distributed computing scheduling environment is deemed to be highly heterogeneous and with processors of uncertain load information. The scheduling objectives are multiple; specifically the focus is on two objectives: minimizing the makespan, cost and energy consumption. The task is to search the optimal set of this MOO problem under the considered environment.

2.1. Problem Formulation

Collaborative computing task allocation is determining how to distribute the processing tasks to appropriate members, who belong to diverse computing resource alliance [23]. With the help of the task planner, computing tasks will be reasonable decomposition. Then, we match and find suitable resources to optimal select a set of effective solution in many candidate providers. The resource scheduling center is considered to have two pieces of information: a collection of user requests and processor information. Each user request is represented by a directed acyclic graph (DAG), which captures a number of task units involved, each unit's own properties, and the relationships among task units. One essential property of each task unit that we must take into account for assignment is the task type. Figure 1 is a collaborative computing task allocation example, when the system platform receives task requests *TR*, the task planner decompose *TR* into *n* sub tasks. Each sub task can be finished by lots of candidate computing resources. In Figure 1, an example of DAGs, each circular node represents a task unit, which has its task type; each rectangular node represents a resource or service, each directed line between two nodes represents their dependency relationship, and we can add weight to the edges to depict the flow size.

$V=(V_i|1=1:n)$ represents the decomposed task units of each user request, where n is the total number of task units. $T=(T_1|1=1:n)$ denotes the task type of each unit in V, where $T_1\in(1,2,\cdots,T_{\max})$ with $T_{\max}$ indicating the total number of task types. $E(n\times n)$ denotes dependencies between task units in *V*. Let $E_{ij}=1$, if data obtained from V is used by V_j. Otherwise, $E_{ij}=0$. $DIN(n\times n)$ represents each task unit's input data size. $DOUT(n\times n)$ represents each task unit's output data size. $P=(P_i|1=1:p)$ represents a collection of processors, where p is the total number of processors. $TP(T_{\max}\times p)$ denotes the computing power of the processor, where TP_{ik} represents time cost for processor P_k to execute the task unit of type i. TP_k denotes the average power of processor k, whose value can be obtained by calculating the mean of elements in column k of matrix TP. S denotes the memory size of each processor. $EP(T_{\max}\times p)$ denotes the computing energy consumption rate, where EP_{ik} represents the energy consumed on processor P_k by executing task unit of type i per unit time per unit data. DC denotes the bandwidth between processors, where DC_{kl} rep-resents the transferring rate of data from processor P_k to processor P_l. EC denotes the communication energy consumption rate, where EC_{kl} represents the energy consumed by transferring data from processor P_k to processor P_l per unit time per unit data. X denotes the mapping between task units and processors. $X(i)=K$ means that task unit V_i has been assigned to processor P_k to be executed.

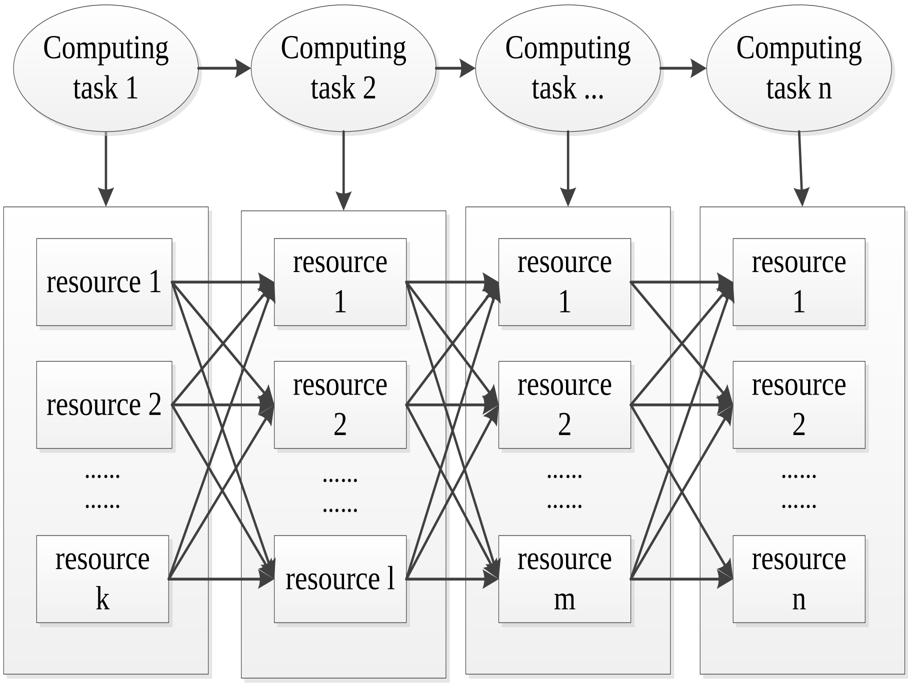

Figure 1. Collaboration Computing Task Scheduling Process

2.2. Objective Functions and Optimization Model

Collaborative computing task allocation has two principles: First, the optional task allocation plan must achieve the shortest time, the lowest cost and highest quality in many service suppliers; second, to solve the problems in the production process, we should make rapid redistribution when business process changes, and finish a project production on time. According to the two principles, this paper establishes a optimization model, taking T (project completion time), Q (quality of computing technology), C (total project cost) as objective function.

(1)The shortest total time T

$$\min T = \sum_{i=1}^{n}\sum_{j=1}^{m_i} x_{ij} t_{ij} \tag{1}$$

In which $x_{ij} \in (0,1)$, denotes that for sub task i, if we choose the j node to complete, then $x_{ij}=1$, otherwise $x_{ij}=0$. t_{ij} represents complete time to the j node, n is the number of all sub tasks.

(2)The lowest cost C

$$\min C = \sum_{i=1}^{n}\sum_{j=1}^{m_i} x_{ij} c_{ij} \tag{2}$$

In which c_{ij} represents complete cost for the j node to finish the task.

(3)The highest quality Q

$$\max Q = \sum_{i=1}^{n}\sum_{j=1}^{m_i} x_{ij} q_{ij} \tag{3}$$

In which q_{ij} represents complete quality for the j node to finish the task.

In the multi-objective optimization process, because each target has different dimension, we need unify objective function index, and transfer multi-objective problem into a single objective by weighted algorithm. In order to make a unified objective function, using the following normalization method:

$$b_{ij} = \left(r_{\max} - r_{ij}\right) / \left(r_{\max} - r_{\min}\right) \tag{4}$$

$$b_{ij} = \left(r_{ij} - r_{\min}\right) / \left(r_{\max} - r_{\min}\right) \tag{5}$$

Finally, the whole objective function is as following:

$$\min f\left(x\right) = w_t T + w_c C + w_q Q \tag{6}$$

In which w_t , w_c , w_q is respectively weight of three targets, and $w_t + w_c + w_q = 1$.

3. Improved Genetic Algorithm

T Genetic Algorithm (GA) is an important branch of evolutionary algorithm commonly used to solve complex computational problems [24]. Its basic idea is to simulate the process of natural evolution: chromosome encoding, genetic manipulation and evolution with by selection, crossover and mutation, a new generation of individuals is generated and evaluated. The fitness values of all individuals can be used to evaluate the rank of each individual and the crowding distance. Then through the trimming operation, the population size maintains stable throughout the evolution process [25]. Considering the complex characters of task in different distributed computing applications, we propose a more comprehensive task scheduling model, collaborative computing task assignment problem is a typical NP-hard problem, and we implement improved GA which is more efficient and effective.

3.1. To Determine the Encoding Method

Floating point encoding method has lots of advantages, such as larger range of representation, high precision, large search space, and is easy to mix with other methods. Due to computational precision, efficiency and convenience, we build our model using floating point decoding method, which resolves our problem of encoding selection.

3.2. The Population Initialization

In the premise of guaranteeing the neural network sensitivity, the net input of each node is near to zero. In order to create more feasible solutions when we optimize FONN weight values using GA, we set the range of the initial population weights [2, -2].

3.3. Fitness Calculation and set other Parameters of GA

GA fitness is very good characterization of individual chromosomes, the larger the fitness is, the better the performance of chromosomes. We apply f(E) =1 / (1+10× E) as fitness function in the model, after FONN training 25times and the output error is E which is the average relative error between the actual output and the expert scoring. Setting population size at 35, terminating conditions are $E \le \varepsilon_1$ or to obtain the maximum number of iterations as 15. We employ selection operation by the roulette wheel method and crossover operation in uniform arithmetic crossover. The crossover probability is 0.6, and the mutation probability is 0.05.

3.4. The process of Evaluation Model

In this paper, we apply the Delphi program to realize the evaluation process, the concrete steps as showed in Figure 2.

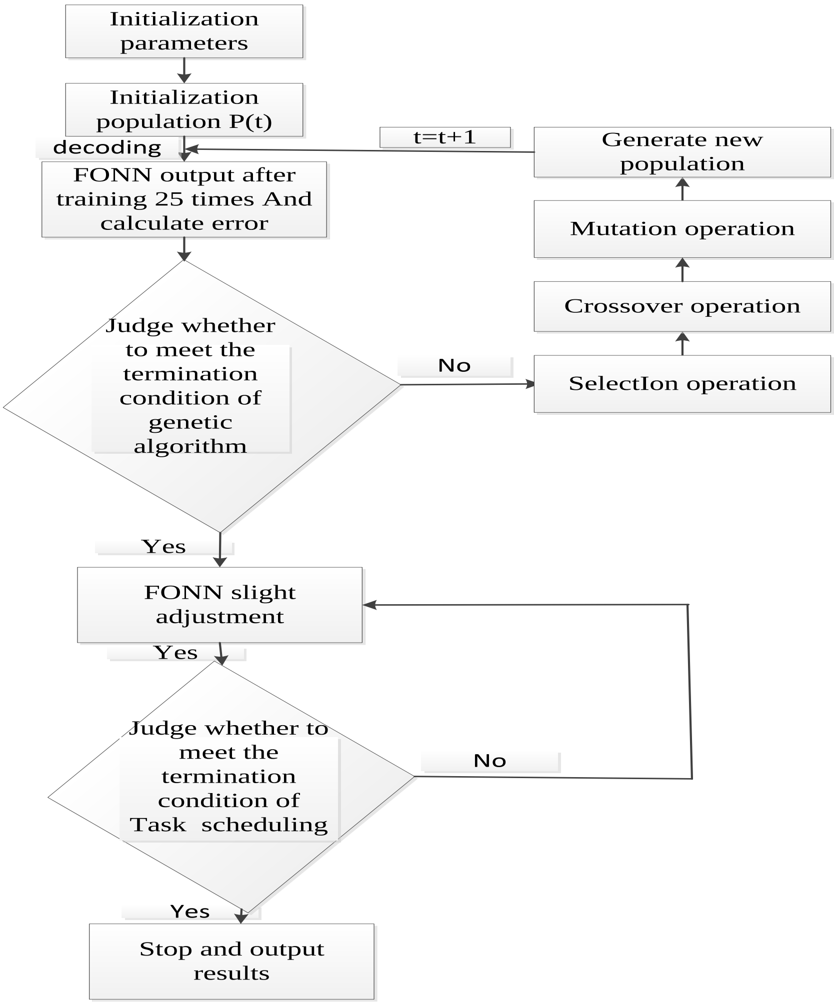

Figure 2. The Process of WSNs Performance Evaluation Model

(1)Firstly, determine the encoding method, terminating conditions, fitness function, other parameters of GA, initialization population size; (2)Decoding to the solution space, after training 25 times, the output error is *E*, then we calculate the individual chromosome fitness $f(E) =1 / (1+10\times E)$;(3)Judge whether to meet the termination conditions of genetic algorithm ($E \le \varepsilon_1$ or to obtain the maximum number of iterations as 15), if to meet, then turn into (5), otherwise we continue step by step; (4) Genetic operation such as selection operation, crossover operation and mutation operation, generate new population and turn to(2); (5) We make slight adjustment of the approximate optimal solutions which is obtained by GA search, improve the accuracy of solutions until it reaches termination conditions, and then output results.

4. A Numerical Example Analysis

In order to test our approach and compare it with other methods, we have built a simulation environment using the C Programming Language. The simulation runs in a step wise manner. In each of iteration, the conditions in the data center evolve with the changing resource requirements of the hosted AEs. The simulator is designed in this way to allow us to compare different methods at identical points in time with the same conditions. The simulation is performed in the laboratory including common software and hardware environment, namely CPU Intel core 4.0GHz, and memory for the DDRII4G, operating system is Windows7.0 professional edition.

In the experiments, the classical roulette wheel selection operator, multiple-point crossover operator and single-point mutation operator are adopted in GA. And the crossover and mutation probabilities are set to be the typical values, i.e. 0.85 and 0.15, respectively. In chaos strategy of IGA and CO, the length of chaotic sequences is set as a constant 10. Due to the randomness of intelligent algorithms, a total of 100 runs of each experiment are conducted and the average fitness value of the best solutions throughout the run is recorded.

4.1. Sample Data

In order to better explain the actual situation, we assume that the distributed computing system gets a computing service request, which contains 7 sub tasks, there are 7 candidate resources to meet each task, and there exists dependent relationship between each task and resource. The complete time, cost, quality of each computing resources can be randomly generated by using simulation platform, then we can collect related available data in Table 1.

Table 1. Statistical Data of Computing Resources

Task number	Resource name	Time	Cost	Quality
1	R_{11}	50	30	9
	R_{12}	45	25	8
	…	…	…	…
	R_{17}	36	75	7
2	R_{21}	55	26	4
	R_{22}	33	54	6
	…	…	…	…
	R_{27}	56	42	6
3	R_{31}	33	74	8
	R_{32}	51	42	6
	…	…	…	…
	R_{37}	76	21	5
…	…	…	…	…
7	R_{71}	45	32	7
	R_{72}	62	24	7
	…	28	62	6
	R_{77}	44	33	7

4.2. Comprehensive Evaluation and Analysis of the Results

In Figure 3, we can see clearly that the run time of the three algorithms all are increasing with service nodes. However, the time of other two algorithms grows exponentially; meanwhile our algorithm is stable, and the run time grows smooth and has smaller fluctuations, which certify our model and algorithm is better and efficient.

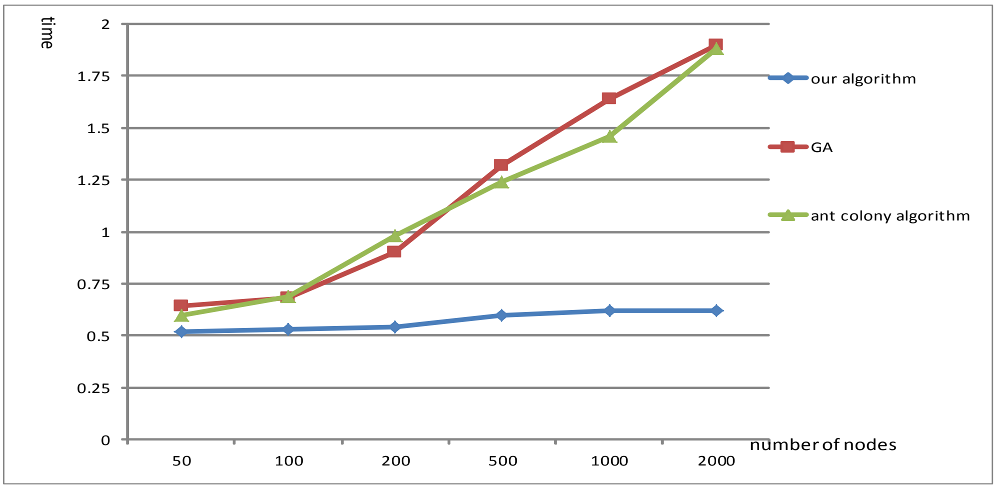

Figure 3. Time Crve with Increase of Service Nodes

As shown in Figure 4, our algorithm tends to be stable after 120 iterations; however, genetic algorithm and ant colony algorithm are iterated 160 times and 200 times. In addition, the average time to search the global optimal solution by the genetic algorithm is 2.4563s, the probability of optimal solution is 0.38; the average time to search the global optimal solution by the ant colony algorithm is 1.8568s, the probability of optimal solution is 0.45. However, the average time to search the global optimal solution by our algorithm is1.2455s, the probability of optimal solution is 0.67. This method is better than the other two algorithms, and we get the best task allocation plan is (R_{12}, R_{23}, R_{37}, R_{45}, R_{53}, R_{66}, R_{71}).

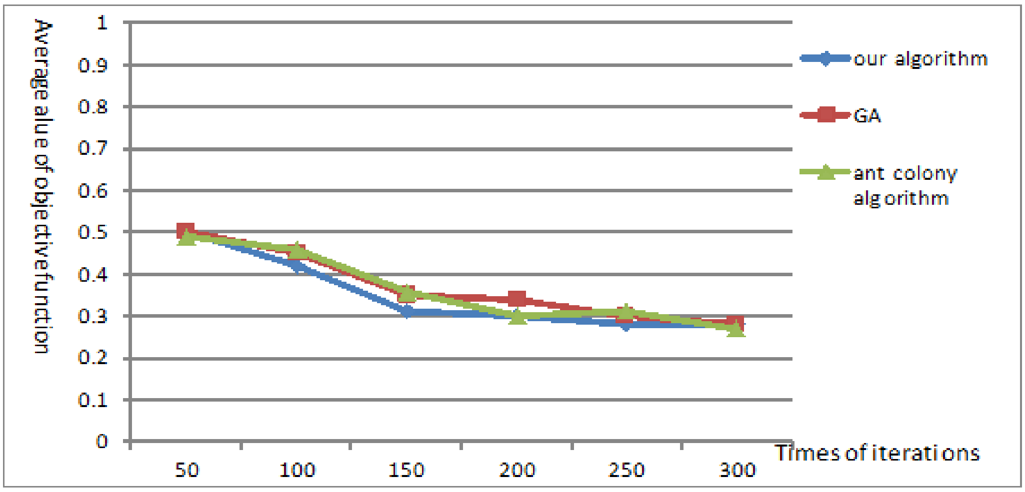

Figure 4. Algorithm Performance Curve

5. Conclusion

In recently years, distributed computing technology has been one of the cutting edge technologies, which has a very vast application prospect and potential utility value. Large scale distributed computing infrastructure are unified computing platform which tries to connect and share all resources in the Internet, including computation resource, storage resource, information resource, knowledge resource and equipment for scientific research, and then solves the problems of large-scale engineering computing and commercial computing as well. Current research of task scheduling distributed computing environments is in its infancy, there is considerable problem space to explore and solve. However, the mutual relations between computing services and the underlying computing resources and the influence of virtualization on quality of services, as two of the key factors in distributed computing system, have not been studied. Meanwhile, previous studies often consider only one objective function, ignoring the complexity and diversity of user' computing task. Therefore, with considering the complex characters of task in different distributed computing applications, we construct a more comprehensive task scheduling model, which has three sub objective functions. And we present an improved genetic algorithm to solve the multi-objective NP-hard problem.

References

[1] T. Ramji, B. Ramkumar and M. S. Manikandan, "Resource and subcarriers allocation for OFDMA based wireless distributed computing system", Advance Computing Conference (IACC), 2014 IEEE International. IEEE, **(2014)**, pp. 338-342.

[2] A. Beloglazov, J. Abawajy and R. Buyya, "Energy-aware resource allocation heuristics for efficient management of data centers for cloud computing", Future Generation Computer Systems, vol. 28, no. 5, **(2012)**, pp. 755-768.

[3] B. J. Lawry, A. Encarnacao and J. R. Hipp, "Node Resource Manager: A Distributed Computing Software Framework Used for Solving Geophysical Problems", AGU Fall Meeting Abstracts, vol. 1, **(2011)**, 1321.

[4] X. J. Chen, J. Zhang and J. H. Li, "Resource reconstruction algorithms for on-demand allocation in virtual computing resource pool", International Journal of Automation and Computing, vol. 9, no. 2, **(2012)**, pp. 142-154.

[5] A. J. Younge, G. Von Laszewski and L. Wang, "Efficient resource management for cloud computing environments", Green Computing Conference, 2010 International, IEEE, **(2010)**, 357-364.

[6] H. Goudarzi and M. Pedram, "Maximizing profit in cloud computing system via resource allocation", Distributed Computing Systems Workshops (ICDCSW), 2011 31st International Conference on. IEEE, **(2011)** January.

[7] P. Pandey and D. Pompili, "Distributed Computing Framework for Underwater Acoustic Sensor Networks", Distributed Computing in Sensor Systems (DCOSS), 2013 IEEE International Conference on. IEEE, **(2013)**, pp. 318-320.

[8] T. Sherif, P. Rioux and M. E. Rousseau, "CBRAIN: a web-based, distributed computing platform for collaborative neuroimaging research", Frontiers in Neuroinformatics, **(2014)**, 8: 54.

[9] Y. C. Lee and A. Y. Zomaya, "Energy conscious scheduling for distributed computing systems under different operating conditions", Parallel and Distributed Systems, IEEE Transactions, vol. 22, no. 8, **(2011)**, pp. 1374-1381.

[10] S. H. Jang, T. Y. Kim and J. K. Kim, "The Study of Genetic Algorithm-based Task Scheduling for Cloud Computing", International Journal of Control & Automation, vol. 5, no. 4, **(2012)**.

[11] P. Varalakshmi, A. Ramaswamy and A. Balasubramanian, "An optimal workflow based scheduling and resource allocation in cloud", Advances in Computing and Communications. Springer Berlin Heidelberg, **(2011)**, pp. 411-420.

[12] F. Xhafa and A. Abraham, "Computational models and heuristic methods for Grid scheduling problems", Future generation computer systems, vol. 26, no. 4, **(2010)**, pp. 608-621.

[13] L. Guo, S. Zhao and S. Shen, "Task Scheduling Optimization in Cloud Computing Based on Heuristic Algorithm", Journal of Networks, vol. 7, no. 3, **(2012)**.

[14] H. Liu H, A. Abraham and A. E. Hassanien, "Scheduling jobs on computational grids using a fuzzy particle swarm optimization algorithm", Future Generation Computer Systems, vol. 26, no. 8, **(2010)**, pp. 1336-1343.

[15] X. Xu, H. Hu and N. Hu, "Cloud task and virtual machine allocation strategy in cloud computing environment", Network Computing and Information Security. Springer Berlin Heidelberg, **(2012)**, pp. 113-120.

[16] J. A. Torkestani, "A new approach to the job scheduling problem in computational grids", Cluster Computing, vol. 15, no. 3, **(2012)**, pp. 201-210.

[17] D. Dutta and R. C. Joshi, "A genetic: algorithm approach to cost-based multi-QoS job scheduling in cloud computing environment", Proceedings of the International Conference & Workshop on Emerging Trends in Technology, ACM, **(2011)**, pp. 422-427.

[18] A. Gorbenko and V. Popov, "Task-resource scheduling problem", International Journal of Automation and Computing, vol. 9, no. 4, **(2012)**, pp. 429-441.

[19] Y. Zhao, L. Chen and Y. Li, "RAS: A Task Scheduling Algorithm Based on Resource Attribute Selection in a Task Scheduling Framework", Internet and Distributed Computing Systems, Springer Berlin Heidelberg, **(2013)**, pp. 106-119.
[20] S. K. Garg, P. Konugurthi and R. Buyya, "A linear programming-driven genetic algorithm for meta-scheduling on utility grids", International Journal of Parallel, Emergent and Distributed Systems, vol. 26, no., 6, **(2011)**, pp. 493-517.
[21] H. Zhong, K. Tao and X. Zhang, "An approach to optimized resource scheduling algorithm for open-source cloud systems", ChinaGrid Conference (ChinaGrid), 2010 Fifth Annual. IEEE, **(2010)**, pp. 124-129.
[22] S. K. Garg, R. Buyya and H. J. Siegel, "Time and cost trade-off management for scheduling parallel applications on utility grids", Future Generation Computer Systems, vol. 26, no. 8, **(2010)**, pp. 1344-1355.
[23] Y. Wen, H. Xu and J. Yang, "A heuristic-based hybrid genetic-variable neighborhood search algorithm for task scheduling in heterogeneous multiprocessor system", Information Sciences, vol. 181, no. 3, **(2011)**, pp. 567-581.
[24] O. L. Sathappan, P. Chitra and P. Venkatesh, "Modified genetic algorithm for multiobjective task scheduling on heterogeneous computing system", International Journal of Information Technology, Communications and Convergence, vol. 1, no. 2, **(2011)**, pp. 146-158.
[25] J. Kołodziej and F. Xhafa, "Enhancing the genetic-based scheduling in computational grids by a structured hierarchical population", Future Generation Computer Systems, vol. 27, no. 8, **(2011)**, pp. 1035-1046.

Authors

Baoli Yuan, he received his Beng in computer application (1998) and MBA (2008) from University. Now he is a full lecturer of Electronic Commerce Department, technology institute. His current research interests include different aspects of Electronic Commerce and Distributed Systems.

Bin Geng, he is a master graduate student, he graduated from Hebei Normal University in human geography professional. Now, he is working in Shijiazhuang vocational technology college, teaching in e-commerce. He is also a lecturer, and his research directions are computer application, e- commerce.

Hongmei Sun, she graduated from Shandong economic college as a master of management in 2005. Now, she works in Shijiazhuang vocational technology college, teaching in e-commerce. She is also a lecturer, and her research directions are computer information management, e- commerce and online marketing.

