

The Optimization of Energy for Cloud Computing

Ji OuYang¹, Caichang Ding^{2,*} and Lu Dai¹

¹Computer College, Dongguan University of Technology, Dongguan, 523808, China; ²School of Computer Science, Yangtze University, Jingzhou, 434023, China

Abstract: The issues include: design and analysis of the key technology of resource allocation in cloud computing environment; combined with time span, load balancing and energy optimization; design of multi-objective optimization model for resource allocation and heuristic resource allocation algorithm; designing the experimental platform of resource allocation algorithm, energy optimization and performance analysis; obtain original achievements in scientific research, for the resource allocation method based on immune algorithm and energy optimization in cloud computing to provide innovative ideas and scientific basis. This research has an important significance for further study on resource allocation and energy optimization in cloud computing environment.

Keywords: Cloud computing, Load balancing, Resource allocation, Algorithm, Innovative, Performance.

1. INTRODUCTION

With information technology being widely applied in all fields recently, the demand of large scale low cost computing capability is becoming more and more urgent. Meanwhile, the continuous growth of network bandwidth making access to resources in remote computers is increasingly matured. Cloud computing was introduced under this setting. With the further development of distributed computing, grid computing and utility computing, cloud computing has become an emerging mode of computing which takes network as a carrier and dynamically allocate resources according to users' requirements on the basis of virtualization technology [1]. The development of computer and the Internet opens up people's field of vision, that is the pace of digitalizing. The digital age has come into our lives. The e-commerce platform in an imperceptible way has become a part of our lives. Online shopping through the network not only saves consumers' and merchants' time, but also helps analyze the demand of online shoppers, to adjust the amount and the type of product. At the same time, the user's experience gradually strengthens, as a large amount of data is produced by the user, and the traditional solution is to upgrade the hardware platform, but with unlimited growth of user data, the user requires higher speed for updating the hardware. In response to this challenge, big data was introduced, which has already been in use since 2014.

Resource allocation is a key technology of cloud computing. At present, the increase in resource utilization rate both saves energy and reduces emission of CO₂ such that it contributes to the reduction of greenhouse effect, being significant in the protection of the whole ecological environment and conforms to the present idea of "Green IT" [2, 3]. Now it urgently needs to develop studies on optimization of energy

consumption of cloud computing resource allocation and its practical application has achieved innovative results with independent intellectual property discussed in the studies of optimization of energy consumption and prototype systems as well as laid the foundation for participating in international and corresponding domestic standards. Therefore, it has very important theoretical significance which is why resource allocation strategy and methods for optimizing energy consumption under the environment of cloud computing are studied. With respect to optimization of energy consumption of resource allocation under the environment of cloud computing, this paper plans to propose a kind of efficient algorithm for resource allocation, establishes resource allocation under the environment of cloud computing into an optimization model with multiple purposes and solves this model with immune algorithm. As a kind of bionic intelligent algorithm inspired by information processing methods of biological immune system, the Immune Algorithm (IA) simulates biological evolution and inheritance processes following biosystem's principle of "struggle for existence and survival of the fittest" and it has become a rising hotspot of study in computer science [4]. While preserving several features of biological immune system such as good diversity, robustness and implicit parallelism, IA has own features and advantages over other inspired optimization algorithms. With self-adaptability and distributive nature and being very suitable for transplanting to cloud computing platform, IA produces offspring on the basis of the optimal individual of parent and substitutes the group convergence with the convergence of the optimal individual [5].

2. OBJECTIVE OF THE STUDY

2.1. Status of Chinese and Overseas Studies

The study of resource allocation under the environment of cloud computing mainly includes three aspects: the first is realizing the balance of resource requirement and allocation with modes of market mechanism; the second is how to real-

*Address correspondence to this author at the School of Computer Science, Yangtze University, Jingzhou, 434023, China; E-mail: dail@dgut.edu.cn

ize the purpose of saving energy by energy-efficient operation of data center and increase resource utilization and; the third is how to realize load balance of system resources and minimization of the task's execution time.

Rajkumar B. *et al.* first proposed the market oriented framework and allocation mode for cloud computing resources based on such framework [6]. The authors just briefed the relevant frameworks of comparing multiple kinds of cloud computing backgrounds and elaborated all kinds of roles existing in the entire cloud computing market, but they did not propose a detailed scheme for the mode of resource allocation. Xindong You *et al.* proposed the market mechanism-based resource allocation strategy [7] to seek the balance point between requirement and allocation of resources by utilizing the market mechanism. In order to solve the dynamically fluctuant resource requirement on cloud computing market, literature [8] proposes a market oriented method for resource allocation. According to the type of each virtual machine, this method employs a model of predictive control system to adjust the corresponding volume for matching different requirements so as to maximize the supplier's income and customer's satisfaction as well as reduce the energy cost and the average latency time of request. Gréhant X. *et al.* proposed a kind of symmetric mapping resource allocation model [9] which is based on cloud computing market and prevents the conflict of interest between the users and the service providers by adding an intermediary main body between the computing tasks and the resources.

Energy-efficient operation of data center is an emerging subject in the field of resource allocation. The virtualization technology has become a main driving force of modern data center. Moreover, how to bring the advantages of virtualization technology into full play for maximizing the utilization of hardware resources and realize the purpose of energy saving, researchers proposed various solutions. On the basis of average response time and average power, Tian Yiming *et al.* set up an energy consumption model of cloud computing system and designed a kind of scheduling algorithm of minimum expected execution energy consumption [2] that meets the performance constraint. The algorithm for dynamically allocating virtual machines has proved that energy consumption can be reduced by migrating virtual machines and closing idle servers [10]. Hsu C.H. *et al.* modeled the problem of virtual resources allocation into the problem of multidimensional knapsack and found the optimal point of energy consumption that the server executes, but they did not propose effective allocation strategy for virtual resources [11].

With respect to the design of resource allocation algorithm, Chenn-Jung Huang *et al.* proposed a genetic algorithm based on self-adapting cloud computing resource allocation mechanism [12]. Anton *et al.* proposed a kind of resource allocation algorithm which is based on energy-aware cloud computing [13]. Guzek *et al.* set the cost function according to energy consumption and time delay and searched for a better virtual resource allocation plan by using heuristic algorithm [14]. Guzek *et al.* designed a kind of energy aware optimization model and resource allocation strategy which is based on genetic algorithm. Chen Zhen proposed a kind of

cloud computing resource allocation algorithm which is based on ant colony optimization algorithm. The test proves that this algorithm is more efficient than other cloud computing resource allocation methods, but this algorithm is only influenced by bandwidth occupation, network load and response time on allocation of cloud resources. In order to realize balanced load of resource allocation, Brototi M. *et al.* proposed a threshold value based on dynamic resource allocation algorithm [15] under cloud computing environment, but they did not take full account of the problem of energy consumption optimization in resource allocation.

Though Chinese and overseas scholars carried out effective studies on the above fields, the problem of resource allocation under cloud computing environment is still facing the challenges as shown below:

- (1) As a key technology of cloud computing, the problem of resource allocation is still in the phase of study. With respect to key technologies that have an influence on resource allocation, mutual coordination and interdependence between resource allocation and task scheduling, there is still no unified industrial standard formed yet.
- (2) The energy consumption in the resource allocation process under cloud computing environment is influenced by many factors, so it is crucial to establish a mathematical model and carry out intensive study on its performance for essentially uncovering the relationship among the dynamic energy consumption of CPU and resource computing capacity, the supply voltage strategy of resources, the amount of resource, supply voltage and CPU frequencies as well as key factors that have an influence on the system's performance.
- (3) It needs further study to design a scientific and reasonable heuristic resource allocation algorithm through intelligent optimization algorithms (such as immune algorithm) according to the customer's requirement and the load of the present node of resource under the premise of meeting load balance and energy consumption minimization of cloud computing system.

2.2. The Content of Study

This paper aims at systematically and intensively studying the problem of resource allocation which is based on energy consumption optimization and immune algorithm under cloud computing environment. The main objectives of the study include:

Designing an immune algorithm based on heuristic resource allocation algorithm according to the user's requirement and loading condition of the present node of resource and under the premise of meeting computer system's load balance, time span and minimization of energy consumption.

This paper fully considers the features of resource allocation under cloud computing environment and new features brought by the optimization of energy consumption. It plans to establish a resource load balance model, dynamic power consumption model and energy consumption optimization model using the hierarchical structure modeling method and a resource allocation optimization model through immune algorithm to transform the problem of cloud computing re-

source allocation into the problem of multi-objective optimization. By quantitatively analyzing their internal relations, it tries to decompose such problems using several key factors. And then, it studies the mathematical expression between these factors and relevant performances of cloud computing and establishes the performance analysis model.

With respect to the above objectives of the study, this paper plans to carry out intensive analysis of key technologies for resource allocation under the cloud computing environment, coordination and dependence between resource allocation and task scheduling, and multi-objective constraint conditions of time span and load balance. The paper also studies the relationship and quantitative expression of the relation among the dynamic energy consumption of CPU and resource computing capacity, the supply voltage strategy of resources, the amount of resource, supply voltage and CPU frequencies, establishes energy consumption optimization model, and then proposes corresponding solutions for load balance, time span and minimization of energy consumption of cloud computing system, and finally verifies the effectiveness of key technologies that are proposed. Details of the relationship among objects, contents and objectives of the study are shown in Fig. (1).

Aiming at different features of study objects, three of the above studies realize corresponding objectives and the contents of these studies are associated to one another and organically integrated.

It has been made clear in section 1.2 that the purpose of resource allocation under present cloud computing environment is to satisfy SLA (service level agreement) by providing efficient execution, but it does not focus on realizing

minimization of energy consumption to arrange resource nodes. In order to explore the optimization of execution efficiency and energy consumption, it has to dispose three key problems:

- (1) The resources shall be described by integrating factors of two aspects that are: computing capability and supply voltage strategy. The computing capability of resource can be expressed with MIPS. MIPS are millions of instructions that can be executed by the resource per second. Supply voltage policy can be described through DVS (dynamic voltage scaling) technology.
- (2) The description of task is taken as the average reaching quality and computing amount within unit time. The computing amount can be described through total number of tasks.
- (3) It shall take full account of load completion ratio of the resource and its value takes the proportion of the amount of the completed tasks to the amount of the assigned tasks.

3. ANALYSIS AND MODELING OF KEY TECHNOLOGIES

3.1. Establishing the Key Technology Analysis Model For Cloud Computing Resource Allocation

Integrating the analysis and description of section 2.2, the following technical route and study methods will be adopted for intensively developing the design of immune algorithm based on heuristic resource allocation algorithm:

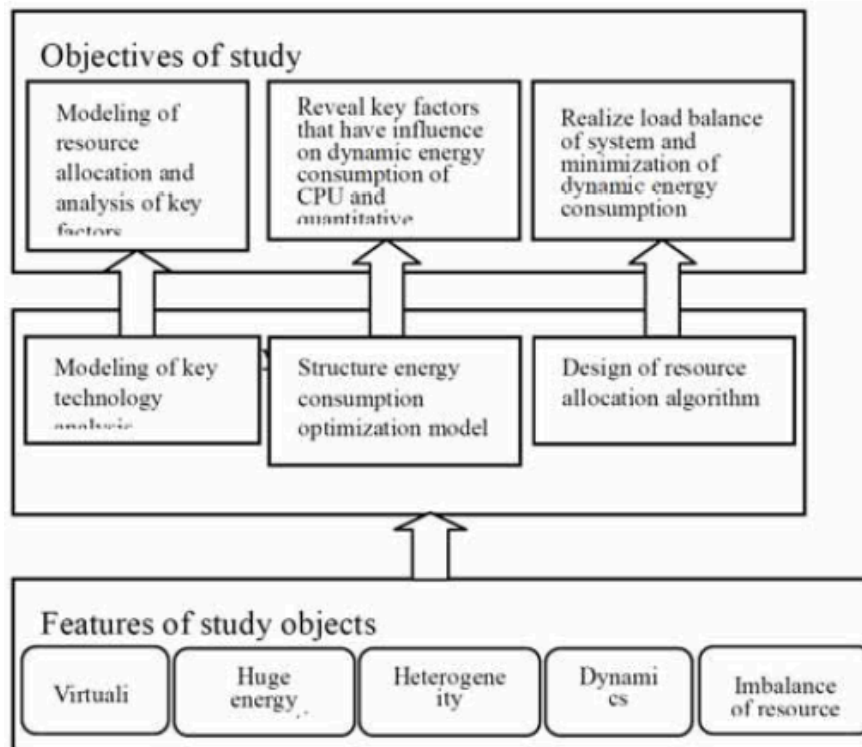


Fig. (1). The relationship among objects, contents and objectives of study.

(1) Structuring multi-objective utility function.

Integrating elements including total energy consumption, resource load balancing factor and time span that are generated by all resources under cloud computing environment to design a multi-objective utility factor and introduce an immune algorithm based on resource allocation algorithm to realize the improvement in the effectiveness of all resources.

(2) Designing partition of equivalent interval.

In order to effectively solve the redundancy problem in immune algorithm, partition of equivalent interval to original population and Logistic chaotic mutation operation is taken into account. The partition of an equivalent interval is an attempt to partition redundant anti-body into one and the same interval block, and then to allocate different operations to different interval blocks and effectively guide the relevant operators of algorithm to carry out instructional search.

(3) By structuring the model of feedback revolution.

In order to strengthen the stability of immune algorithm, this paper introduces an in-depth model of feedback revolution. The preliminary frame of this model is shown Fig. (3), where A represents the open-loop amplification operation and it includes operations of colon, variation and selection in traditional immune algorithm; F represents feedback revolution depth and feedback anti-bodies with lower survivability to feedback population $A_f(t)$. $A_d(t)$ is the biased population which eliminates antibodies with inferior performance from the revolutionary population $A(t)$ (antibodies in $A_f(t)$). At this time, only antibodies with larger antigen affinity and lower concentration are preserved and other antibodies are eliminated. Through this kind of inhibiting effect that is similar to cells, the diversity of antibody population is preserved so as to effectively strengthen the global searching capacity and multi-solution searching capability of algorithm.

(4) Theoretical analysis and test.

The performance of the proposed method is verified by the mode of integrating theoretical, simulation and test analysis, and the immune algorithm based on resource allocation algorithm is not improved and optimized by integrating theoretical, simulation and test data until the optimal performance is achieved.

Integrating the analysis and description of section 2.2.1, the following technical route and study methods will be taken for intensively developing the key technologies analysis model of cloud computing resource allocation:

(1) Description of properties of resource and performance analysis.

Defining the resource r_i as two-tuples $r_i = \{c_i, s^i\}$, where c_i represents the computing capability of resource r_i in unit time, e.g. the number of instructions executed by resource within unit time; s^i represents the supply voltage strategy of resource r_i and the supply voltage of each strategy is different from one another. The main parameters for measuring the computing capability of resource node include the amount of CPU m , the processing capability p (MIPS) and bandwidth B of its network (it determines the time when the task is allocated till it is transferred to the

source node). Meanwhile, it also has to introduce a load balancing factor LB , i.e. the load completion ratio of resource, $LB = La / Lf$, where La represents the amount of completed task, Lf represents the summation of the allocated tasks' computing amount. If the supply voltage strategy s^i has k levels of DVS, then the relation matrix between supply voltage and frequency is described as:

$$V_i = [(v_1(i), f_1(i)); (v_2(i), f_2(i)); \dots; (v_k(i), f_k(i))]^T,$$

Where $v_k(i)$ represents the supply voltage of resource at DVS level k , $f_k(i)$ represents the operating frequency at this time and its range is [1, 2].

(2) Property description and performance analysis of task.

Tasks that randomly reach to cloud computing system are described as set,

$$T = \{t_1, t_2, \dots, t_j, \dots, t_n\}, t_j$$

represents the j th task. t_j is defined as two-tuples

$$t_j = \{a_j, w_j\},$$

Where a_j represents the average reaching amount of task within unit time, w_j represents the computing amount or work load of task t_j . According to the computing amount of the task and the computing capability of resource, the time $ETC(i, j)$ in which the resource r_i executes the task t_j is:

$$ETC(i, j) = \frac{w_j}{c_i} \tag{1}$$

According to this, the computing time increases with the reduction of supply voltage and frequency, the execution time matrix of the task t_j on resource r_i whose supply voltage strategy is s^i is:

$$ETC[i, j] = [\frac{1}{f_1(i)} \times ETC(i, j), \frac{1}{f_2(i)} \times ETC(i, j), \dots, \frac{1}{f_k(i)} \times ETC(i, j)] \tag{2}$$

(3) Theoretical analysis and test.

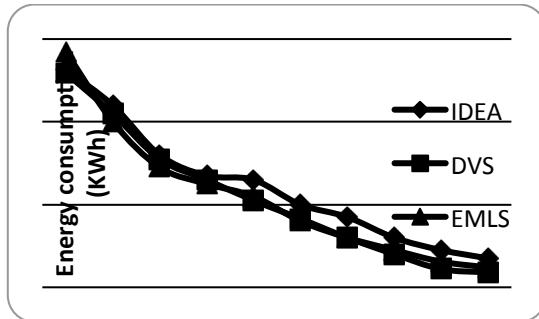
The performance of proposed method is verified by the mode of integrating theoretical, simulation and test analysis and further optimized on the basis of analysis simulation and test data for providing high resource utilization rate and energy-efficient data center and laying basis for design of allocation of resources at upper level.

4. SIMULATION EXPERIMENT

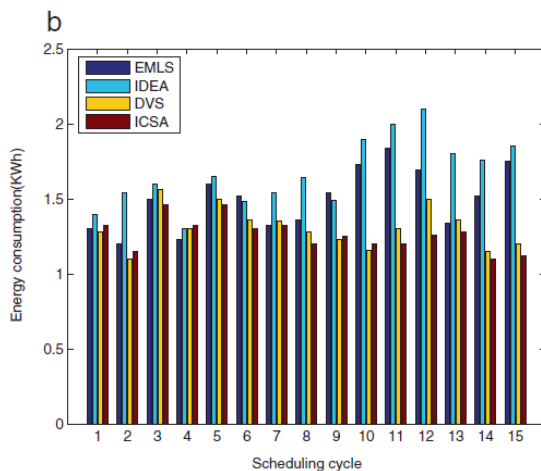
In this section, we analyze the performance of our algorithm based on the experimental results. In order to make it easier to test the algorithms, CloudSim [16] has been adopted in this work as an effective cloud computing simulation platform. Six physical machines have 8 GB RAM and 2 TB storage space, and each machine has four CPUs that have a capacity power of 10,000 MIPS. A data center with 16 DVS-enabled processors was used.

For testing the effectiveness and superiority of the resource allocation model in cloud computing, the existing resource allocation approaches such as the IDEA [17], EMLS [18], and DVS [19] were used. Y-axis represents the

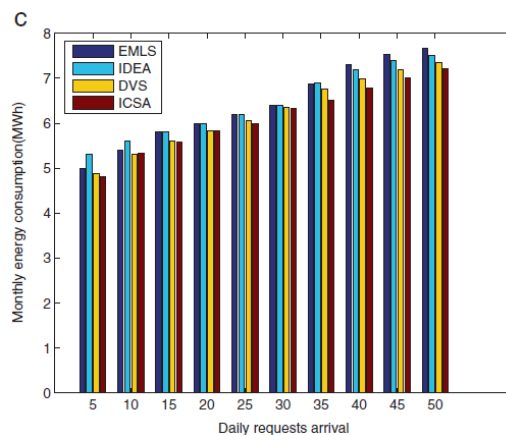
makespan, and X-axis denotes generations and the number of tasks, respectively. Finally, we compare the energy consumption of the three resource allocation algorithms in Fig. (2). In Fig. (2), Y-axis shows energy consumption, and X-axis denotes lower utilization threshold. Response time is the amount of time taken by the submission of a request and the first response that is produced [13, 14].



(a) Number of tasks



(b) Generations



(c) Generations

Fig. (2). Comparison of the makespan of the four resource allocation algorithms. (a) Different tasks. (b) $n = 400$. (c) $n = 600$. (d) $n = 800$.

Energy efficiency is one of the key technologies of resource allocation in cloud computing. When the lower utilization threshold increases, the energy consumption of the system is also rapidly reduced. The statistical analysis of the

monthly energy consumption comparison of the four algorithms is illustrated. It can be observed that there is a significant difference among the four resource allocation algorithms, and our proposed ICSA consumes the least energy in most cases. Through the above experimental results, it can be observed that ICSA can effectively meet the requirements of resource and can save much more time compared to the other approaches. The ICSA is well enhanced and balanced on exploration and exploitation and has better stability and scalability. Thus, the ICSA shows its effectiveness to improve energy efficiency of the data center and decrease its makespan.

It can be observed that there is a significant difference among the three resource allocation algorithms. The DVS is well enhanced and balanced in terms of exploration and exploitation and has better stability and scalability.

CONCLUSION

Green cloud computing is the future development trend and the main research object. Reducing energy consumption is an increasingly important issue in cloud computing, more specifically when dealing with a large-scale cloud. In this paper, we proposed an improved clonal selection algorithm based on time, cost and energy consumption models in cloud computing environment. The experimental results show that our approach has immense potential as it offers a significant improvement in average execution time, demonstrates high potential in improving energy efficiency of the data center, and can effectively meet the service level agreement requested by the users. It studies the resource allocation under cloud computing environment by integrating the immune algorithm, energy consumption design and optimization technology. The study is mainly based on the knowledge and technologies in the fields of cloud computing, electrical engineering and intelligent computing. There have been many basic theories and research achievements in such fields, which lay sound theoretical basis and provide referential technologies for the study to be carried on under this subject.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

Declared none.

REFERENCES

- [1] J.T. Tsai, J. C. Fang, and J.H. Chou, "Optimized task scheduling and resource allocation on cloud computing environment using improved differential evolution algorithm," *Comput. Oper. Res.*, vol. 40, no. 2, pp. 3045-3055, 2013.
- [2] Y. Kessaci, N. Melab, and E.G. Talbi, "A multi-start local search heuristic for an energy efficient VMs assignment on top of the OpenNebula cloud manager," *Futur. Gener. Comput. Syst.*, vol. 29, no.1, pp. 1-20, 2013.
- [3] J. Kołodziej, S.U. Khan, L. Wang, M. K. Dorohinicki, and S.A. Madani, "E Niewiadomska-Szynkiewicz, AY Zomaya, Security, energy, and performanceaware resource allocation mechanisms for computational grids," *Futur. Gener. Comput. Syst.*, vol. 29, no.1, pp. 944-959, 2013.

- [4] S. Khan, "A goal programming approach for the joint optimization of energy consumption and response time in computational grids," in Proceedings of the 28th IEEE International Performance Computing and Communications Conference, pp. 410–417, Piscataway, IEEE, 2009.
- [5] S.U. Khan, and I. Ahmad, "A cooperative game theoretical technique for joint optimization of energy consumption and response time in computational grids," *IEEE Trans Parallel Distr. Syst.*, vol.20, no.3, pp.346–360, 2009.
- [6] Y. Wang, T. Yang, Y. Ma, G.V. Halade, J. Zhang, M.L. Lindsey, and Y.F. Jin, "Mathematical modeling and stability analysis of macrophage activation in left ventricular remodeling post-myocardial infarction," *BMC Genomics*, vol. 13, pp. 1–8, 2012.
- [7] A. Merkel, and F. Bellosa, "Memory-aware scheduling for energy efficiency on multicore processors," In: *Proceedings of the Workshop on Power Aware Computing and Systems*, pp. 123-130, Pearson Prentice Hall, Englewood Cliffs, 2008.
- [8] G.C. Jr., G.C. Nakamura, D.G. Chavarría and H.J. Sofia, "Graph Mining of Networks from Genome Biology", In: *Proceedings of the 7th IEEE International Conference on Bioinformatics and Bioengineering, BIBE*, 2007.
- [9] D. A. Keim, "Information visualization and visual data mining", *IEEE Transactions on Visualization and Computer Graphics* vol. 8, no.1, 2002.
- [10] V. Niennattrakul, D. Wanichsan, and C.A. Ratanamahatana, "Accurate subsequence matching on data stream under time warping distance", *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, pp. 5669, 2010.
- [11] P.G. Ferreira, and P.J. Azevedo, "Protein sequence pattern mining with constraints", *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, pp. 3721, 2005.
- [12] G. Ivosev, L. Burton, and R. Bonner, "Dimensionality reduction and visualization in principal component analysis," *Anal Chem*, vol. 80, 2008.
- [13] M. Rahman, T.L. Haque and T. Fukui, "Research articles published in clinical radiology journals: Trend of contribution from different countries", *Academic Radiology*, vol.12, 2005.
- [14] S. Curteanu, F. Leon, and D. Galea, "Alternatives for multiobjective optimization of a polymerization process," *Journal Applied Polymer Science*, 2006.
- [15] B. Mondal, K. Dasgupta, and P. Dutta, "Load Balancing in Cloud Computing using Stochastic Hill Climbing-A Soft Computing Approach," *Procedia Technology*, vol. 4, no. 5, pp.783-789, 2012.
- [16] W. Lin, J.Z. Wang, C. Liang, and D. Qi, "A Threshold-based Dynamic Resource Allocation Scheme for Cloud Computing", *Procedia Engineering*, vol. 23, no. 11, pp. 695-703,
- [17] W. Shu, S. Zheng, L. Gao, et al, "An Improved Genetic Simulated Annealing Algorithm Applied to Task Scheduling in Grid Computing," *Dynamics of Continuous, Discrete & Impulsive Systems-Series B: Applications & Algorithms*, vol. 13, no. 2, pp. 831-835, 2006.
- [18] W. Shu, L. Ding, and S. Wang, "Resource Sharing Models and Heuristic Load Balancing Methods for Grid Scheduling Problems," *International Journal of Advancements in Computing Technology*, vol. 4, no. 9, pp. 315-322, 2012.

Received: September 16, 2014

Revised: December 22, 2014

Accepted: March 03, 2015

© OuYang et al.; Licensee Bentham Open.

This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.