

Based on Particle Swarm Optimization Algorithm of Cloud Computing Resource Scheduling in Mobile Internet

Yong Lin

Qingdao Vocational and Technical College of Hotel Management
214300654@qq.com

Abstract

Mobile Internet due to the limitation of the mobile terminal power supply, transmission and calculation of the need to adopt energy saving strategy, also due to the terminal mobility, mobile Internet topology change. Therefore, mobile Internet cloud computing resources allocation need both energy efficiency and assure the time characteristics of mobile business; Aiming at this problem, this paper presents a maximum energy efficiency optimization, in the restrictive conditions at the same time, guarantee the minimum time delay the business. According to the characteristics of the optimization problems, both the distribution of the improved algorithm is proposed, the algorithm based on particle swarm optimization (psa) algorithm, build the search direction matrix of orientation, the simulation results show that the proposed allocation algorithm can effectively improve the efficiency of energy utilization, and ensure that the time delay of the business requirements.

Keywords: *The Mobile Internet, Cloud Computing, Particle Swarm Optimization Algorithm, Resource Scheduling*

1. Introduction

With the development of computing and communication capabilities of mobile devices, while the storage capacity has been greatly improved, such as intelligent mobile phone, laptop computer and tablet computer *etc.* Statistics show that, by 2015, mobile devices will exceed wired equipment; they have become the main equipment to access the Internet. The change from the terminal access problems affect the distributed computing resource pool, the mobile Internet access network and the core of the mobile Internet and its computing resources and storage resources and the number of nodes are different from the traditional Internet, which requires resource scheduling in different ways, at the same time, as the business application of Intelligent Mobile phone and tablet computer *etc.*, there are many sensing devices and sensor data, such as gravity, direction and GPS positioning information, can provide the spatial distribution information, thus bringing convenience to resource scheduling. At the same time, cloud computing can make mobile Internet access to the following three aspects: first, resources can be stored in the network resources; second, computing resources can be obtained in the network; third, application service resources of other access nodes can be obtained in the network.

Although cloud computing can be used to compute and store resources allocation, but the traditional Internet cloud computing have not considered the problem of bandwidth and energy consumption, so the traditional Internet cloud computing methods cannot be directly applied to the mobile Internet environment. Mobile Internet and cloud computing is getting more and more attention from the academic community,

Mobile Internet mainly in the mobile terminal wireless access equipment, and the energy is limited, in the fifth generation mobile communication system, the concept

of green communication to win support among the people, but the design is mainly for data transmission in physical layer and resource scheduling in data link layer, which is mainly concentrated on the data transmission in network; however, through the application layer, there is little research involved that reasonable scheduling of mobile internet terminal computing and storage resources methods.

[4] applied a principle of continuous time Lagrange dual to schedule data resources, the objective function of optimization is to minimize transmission energy consumption. [5] studied online perceived efficiency and transmission delay structure. [6] considered the time-varying channel conditions control algorithm based on energy efficiency. However, the conditions of the design which above mentioned algorithm is too ideal, considering the transmission of data to only one user.

Then [7] a base station in N time-varying channels for N users to provide services, and proposed an adaptive channel access algorithm. [8] proposed the method based on adaptive carrier aggregation to change data download and upload efficiency. As can be seen, the existing researches are based on transmission of the physical layer and scheduling combination of the application layer, the focus remains on the physical layer, the application layer of the business is mainly concentrated in the upload and download data like this, this is obviously to find that it did not meet the need of mobile Internet. [9] studied on the work calculation model of Internet, according to cloud computing in the online computation with dynamic allocation. In order to improve the efficiency of distribution, [10] used a competitive bidding will work out the distribution load calculation. This method simplifies the calculation complexity of the distribution algorithm, but its effect is not as good as the other distribution algorithm which has high computational complexity. On the basis of this, [11] proposed resource allocation that joint optimize the demand and distribution, this way can fast allocate resource, and it has low computational complexity. But these algorithms cannot be applied to the mobile Internet, because its computation capability of single node is low, and the communication ability between each node is lower than that of the internet. We need to find a suitable scheduling algorithm in mobile internet.

According to the research state above, the main goal of this paper is to optimize the mobile Internet energy efficiency, the computing resources of mobile internet need to be allocated. Because a lot of business is offered by mobile Internet, so we cannot simply use the data transmission rate to measure. To solve this problem, this model does not consider the specific number of terminals, so the mobile Internet application scenarios are more robust. Although this is only consider energy efficiency problems of mobile Internet computing resource allocation, but the model provides an important reference for the follow-up study. In a follow-up study, it needs to consider the delay, and BER performance that more strict conditions, in order to adapt to the real needs that the mobile Internet applications business is various.

2. Problem Descriptions

In the mobile Internet, scheduling network model different from the traditional Internet network model, allocating and scheduling users in the cell used the way of stratified the base station. All users in the cell can be scheduled and assigned tasks, and can also get service from other users. The specific structure shown in Figure 1:

In the mobile Internet, setting n type of business needs to adjust, because there are many of the mobile Internet business types, so the n is great, it provides business are S_1, L, S_n , and calculation resources is required for each business type is r_i , in this paper, using the vector $r = (r_1, r_2, L, r_n)$ represents calculation resources of n

different business types. Setting of computing resources in the mobile Internet can be mobilized divided into m classes, each class with calculation function for c_i , and c_1 is the normalized calculating function, so $c_1=1$; computing ability of the virtual machine can be expressed by the vector $C=(c_1,c_2,\dots,c_m)$, as c_1 is the normalized type, so the vector C can be regarded as the proportional numbers of calculation ability. For example, if the virtual machine have 4 kinds of computing power can be provided by cloud computing service provider of mobile Internet, respectively, c_1, c_2, c_3 and c_4 , where, computing resources can be mobilized by c_1 , processing frequency is 1GHz, 2GB of call memory and 500GB of hard disk; at the same time, c_2 has dual core the computational resources and processing frequency 1GHz, contains 4GB of call memory and 1TB of hard disk ; c_3 has quad-core the computational resources and 1GHz of processing frequency, 8GB of call memory and 2TB of hard disk ; c_4 has magny-cours of computational resources and 1GHz of processing frequency, 16GB of call memory and 4TB of hard disk ; It is composed of

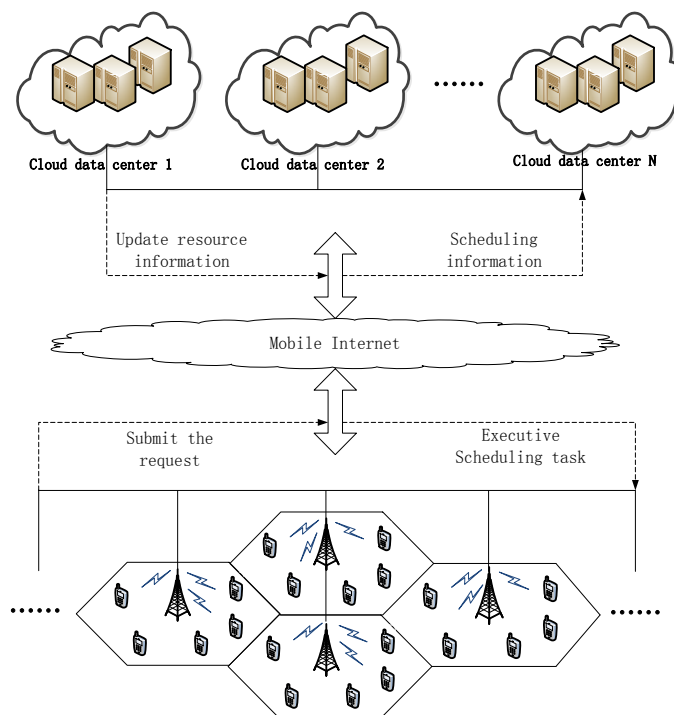


Figure 1. Mobile Cloud Computing Scheduling Model

ability of computing resources for vector $C=(1,2,3,4)$. If the mobile Internet cloud computing service can provide the number of the i -th class is l_i , while vectors consists of the number of all kinds of virtual machines provided is $L=(l_1,l_2,\dots,l_m)^T$, it should meet the conditions for

$$CL \leq M \quad (1)$$

Where, M represents the total number of the various computing services can provide. In this part, mainly consider $T=\{t_1,t_2,\dots,t_n\}$ is the set of all tasks to handle; exchanging treatment data that need to be processed between t_i and t_j , then exchanging data is d_{ij} , maximum mutual information through the channel can be

expressed as d_{ij} , let total computing ability of i -th the virtual machine be ρ_i , the time needed to complete the task can be expressed as (2)

$$T_{ij} = d_{ij} / (\rho_i * \rho_i) \quad (2)$$

Where, ρ_i represents computing ability of this nodes in this task. But because of bandwidth limitations, it needs to consider the transmission time of node transmission information, let transmission time (including transmission time of before and after treatment) be T_{ct} , so the total completion time for the business is T_{z_i} . According to computing services of the virtual machine provided, supposing the energy consumption of each virtual machine computing service computing unit is E_i , due to transmission system is different, according to the system network, energy of unit data transmission that needed (including the required uplink and downlink data, thus energy) is E_{ct} , so the energy required can be obtained

$$E_{z_i} = E_{ct} d_{ij} + E_i t_1 \quad (3)$$

According to the definition of the above model, the needed resource allocation can be defined, in the past optimization models, the business processing time often was took seriously, namely the delay of processing business, which is also an important indicator of the fifth generation mobile cellular communications, but it has also an important index is handling business required energy consumption, aiming at the problem of energy consumption, this paper put forward the optimization of energy consumption and taking delay into account, so it put forward the following optimization problems

$$\begin{aligned} \min E_Z \\ \text{s.t. } T_{z_i} \leq \eta_i \end{aligned} \quad (4)$$

Where, the consumption energy of all tasks is E_Z , it can be shown by (5)

$$E_Z = \sum_{i=1}^M E_i \quad (5)$$

3. Resource Scheduling Design based on Particle Swarm

In the actual search process, due to the heterogeneous network, the user changes, resulting in changes of the virtual machine, each task in the data transmission energy consumption will change, so computing resource scheduling rate is also the key issues to consider. This paper selected the particle swarm algorithm, this algorithm can guarantee the search solution has the fastest convergence rate, and it can obtain the performance of parallel computing, but due to the solution by particle swarm algorithm searched is easy to fall into the local optimal solution, thus affecting the resource allocation effect, therefore, this paper modified particle swarm algorithm, in order to ensure the searched solution with the characteristics of the optimal solution, and it can ensure the convergence speed.

First, determining the particle is energy consumption of the task, but for the search direction of energy consumption, this paper intends to use the gradient progressive method, namely, according to the direction of the partial derivative to search, *i.e.*

$$\sigma = \frac{\partial E_Z}{\partial E_i} \quad (6)$$

And the basis of this search is the end of the last search, but the initialized starting point, we selected the calculated initial energy consumption as the search starting point. It can search in the following form:

$$E_i(n+1) = E_i(n) + \mathbf{v}_i(n) \quad (7)$$

$$\begin{aligned} \mathbf{v}_i(n+1) = & \sigma \mathbf{v}_i(n) + \lambda_1 \theta_1 (E_i^*(n) - E_i(n)) \\ & + \lambda_2 \theta_2 (E^*(n) - E_i(n)) \end{aligned} \quad (8)$$

Where, σ represents the convergence rate proportion from last search step; λ_1 and λ_2 are the proportion vector of two conditions, this vector was used to search the required direction, the combination is two-dimensional direction adjustment. β_1 and β_2 are the acceleration coefficient, $E_i^*(n)$ on behalf of the optimal solution of the search, the specific location of the search is $E_i(n)$.

According to the design of genetic algorithm, we need to design the search direction, according to the direction to search, but due to the different standards, the search direction of the optimal solution is to determine the orientation matrix, in order to avoid the shortcomings of the traditional genetic algorithm - the situation of a local optimum, the direction of movement of the orientation matrix transformation should be transformed, so as to ensure search the global optimal solution. When setting the global optimal solution, it needed to compare the search target that is closer to the global optimum, which determine the starting point of the next hop, as shown in (9), the matrix form is

$$C = \begin{bmatrix} 0 & L & c_{1n} \\ M & O & M \\ c_{n1} & L & 0 \end{bmatrix} \quad (9)$$

If compared every solution, if solution 1 more close to the global optimal solution (with respect to the solution 2), the matrix value is +1, and if close at the same time, so assigned to 0, if away from the global optimal solution, assigned to -1. Because of the orientation matrix expressed by only three discrete values, the accuracy is not accurate, it can be compared each element further, thus forming the final orientation matrix. Also in this way, the kronckern product form needs to be used, as shown in (10)

$$C_z = \mathbf{C} \otimes \mathbf{C} \quad (10)$$

According to the final orientation matrix C_z , optimized direction can be obtained. And according to the optimization objective function, namely the energy consumption randomly generated initial value. The main consideration is energy consumption and calculation and transmission time. According to the two main performance value allocation related resources, optimization the allocation resources can be achieved.

The energy consumption for search can be analyzed by the following matrix, as shown in (11)

$$E = \begin{bmatrix} e_{11} & e_{12} & L & e_{1n} \\ e_{21} & e_{22} & L & e_{2n} \\ M & M & O & M \\ e_{m1} & e_{m2} & L & e_{mn} \end{bmatrix} \quad (11)$$

The matrix can be used as execution time can use (12) to show, its definition can be expressed as

$$\mathbf{T} = \begin{bmatrix} t_{11} & t_{12} & L & t_{1n} \\ t_{21} & t_{22} & L & t_{2n} \\ M & M & O & M \\ t_{m1} & t_{m2} & L & t_{mn} \end{bmatrix} \quad (12)$$

According to (11) and (12), σ and λ_1, λ_2 the acceleration coefficient can be calculated. These coefficients can according to (13), (14) and (15) shown

$$\beta_1 = \frac{\left(\|(\mathbf{E})_i\|_2^2 - \|(\mathbf{E})_j\|_2^2\right) + \left(\|(\mathbf{T})_i\|_2^2 - \|(\mathbf{T})_j\|_2^2\right)}{\|(\mathbf{E})_j\|_F^2 + \|(\mathbf{T})_j\|_F^2} \quad (13)$$

$$\beta_2 = \frac{\left(\|(\mathbf{T})_i\|_\infty^2 - \|(\mathbf{T})_j\|_\infty^2\right) + \left(\|(\mathbf{E})_i\|_\infty^2 - \|(\mathbf{E})_j\|_\infty^2\right)}{\|(\mathbf{T})_j\|_F^2 + \|(\mathbf{E})_j\|_F^2} \quad (14)$$

$$\beta_2 = \frac{\left(\|(\mathbf{T})_i\|_\infty^2 - \|(\mathbf{T})_j\|_\infty^2\right) + \left(\|(\mathbf{E})_i\|_\infty^2 - \|(\mathbf{E})_j\|_\infty^2\right)}{\|(\mathbf{T})_j\|_F^2 + \|(\mathbf{E})_j\|_F^2} \quad (15)$$

So according to the design parameters above, and the particle swarm algorithm, the resource allocation algorithm can be summarized as follows

improved algorithm configure the virtual machine

Input: $\mathbf{r} = (r_1, r_2, \mathbf{L}, r_n)$ $T = \{t_1, t_2, \mathbf{L}, t_n\}$

Output: $\mathbf{L} = (l_1, l_2, \mathbf{L}, l_m)^T$

- 1: initializing $\mathbf{r} = (r_1, r_2, \mathbf{L}, r_n)$,
 - 2: for $(j=1, j \leq n, j++)$ do
 - 3: collecting $A_j = (r_1^j, \mathbf{L}, r_m^j, t_j, e_j)$
 - 4: end for
 - 5: Setting parameters and orientation matrix C
 - 6: for $(j=1, j \leq n, j++)$ do
 - 7: computing the total amount of request calculate resources to the user u_j
 - 8: computing the maximum energy of unit resources to the user u_j
 - 9: computing processing time to the user u_j
 - 10: end for
 - 11: according to the maximum energy consumption e_j for users to sort, store $e_1 \geq e_2 \geq \mathbf{L} \geq e_n$
 - 12: for $(j=1, j \leq n, j++)$ do
 - 13: if $v_j/s_j \geq v_{re}$ then
 - 14: $l++$
 - 15: end if
 - 16: refused to request of user u_j, \mathbf{L}, u_n
-

```
17:   (11) (12) and (13-15)
18:   end for
19:   for (j=1, j ≤ n, j++) do
20:     if  $b_j \leq B$  then
21:       if  $s_j \leq R$  then
22:          $W = W \cup u_k, R' = R - s_j,$ 
23:       end if
24:     end if
25:   end for
26:   for (i=1, i ≤ m, i++) do
27:      $k_i = \sum_{j:u_j \in W} r_i^j$ 
28:   end for
```

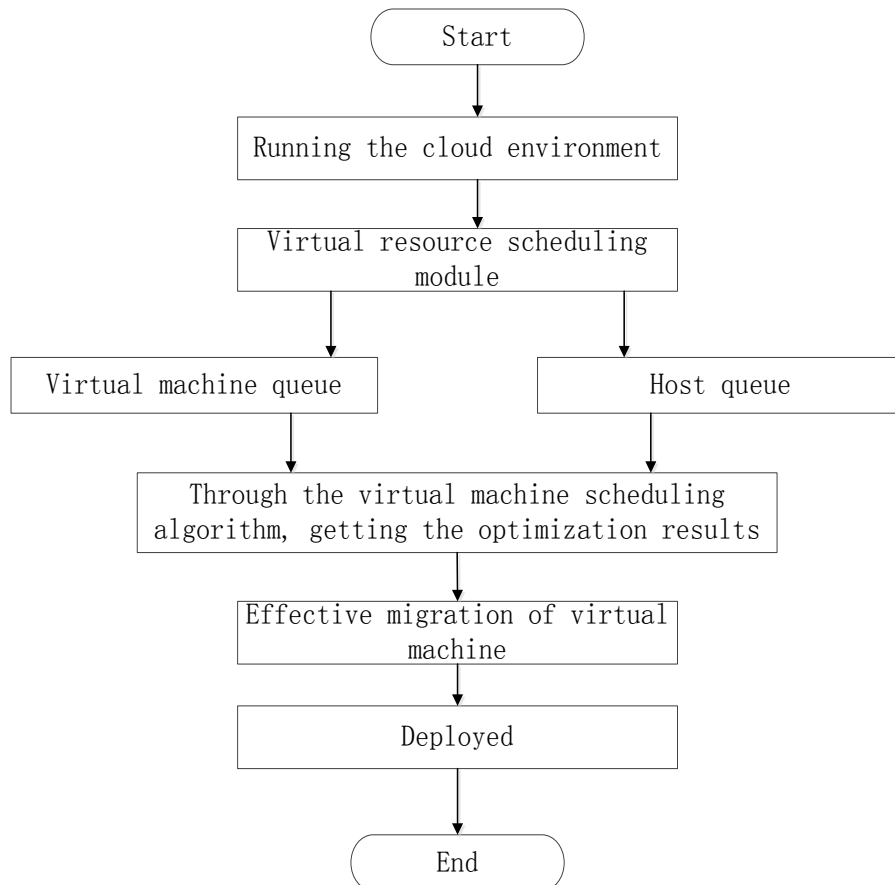


Figure 2. The Resource Allocation Process

4. Experiment Results

In this paper, the calculation and simulation of the cloud computing model for the mobile Internet, so it needs to build a cloud simulation platform, this model platform used cloud computing and simulation common platform CloudSim, it developed on the basis of the distributed parallel calculations, using this platform can through resources of computer to simulate data storage and transfer, but lacking of topology changes link, according to the practical situation, this experiment has been modified, the modified based on the topological graph to modify its transmission data and transmission time. The simulation environment included the configuration of computer environment. The computer simulation environment as shown in Table 1

Table 1. VM Configuration

	processor	memory	hard disk
VM_1	1 × 2 GHz	4 GHz	500 GB
VM_2	2 × 2 GHz	8 GHz	1 TB
VM_3	4 × 2 GHz	16 GHz	2 TB
VM_4	8 × 2 GHz	32 GHz	4 TB

The distance of the simulation is shown below, achieving the model based on virtual tasks and scheduling, the core algorithm for the development of writing, the scheduling interval according to the simulation environment is different; it needed to set up a separate.

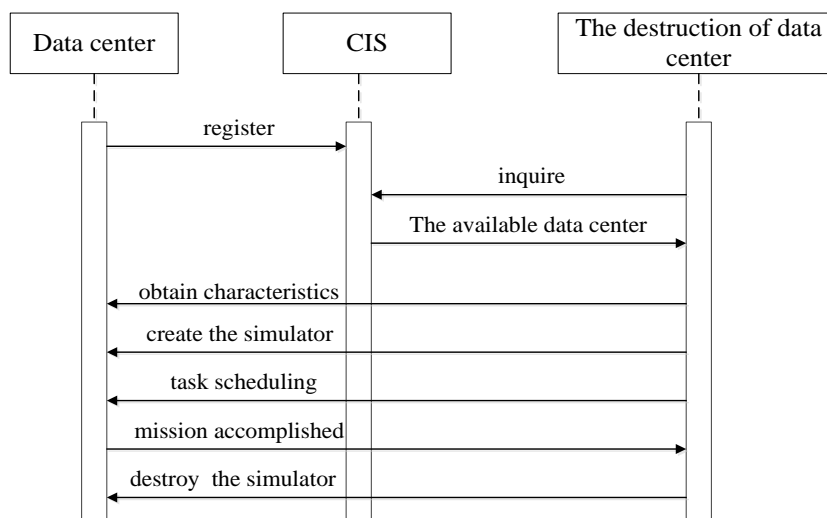


Figure 3. Cloud Computing Simulation Process

Figure 4 shows that the improved scheme proposed in this paper has the advantages which the energy utilization rate, according to the different topology and execution time, the cost is lower than that of the traditional particle swarm optimization scheme. And with the growth of data arrival rate, the advantage is more obvious.

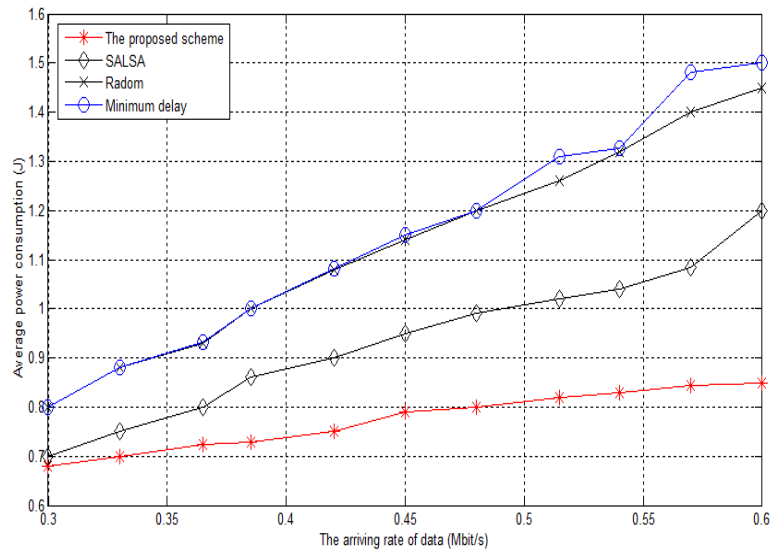


Figure 4. The Average Energy Consumption Comparison

Figure 5 shows resource utilization rate of the proposed algorithm is higher than that of the traditional particle swarm method, this method because of high resource utilization rate, which can give full play to the advantages of distributed computing, thus saving computational cost. And we can see the advantages of the algorithm in terms of time, because the resources utilization rate includes time resources.

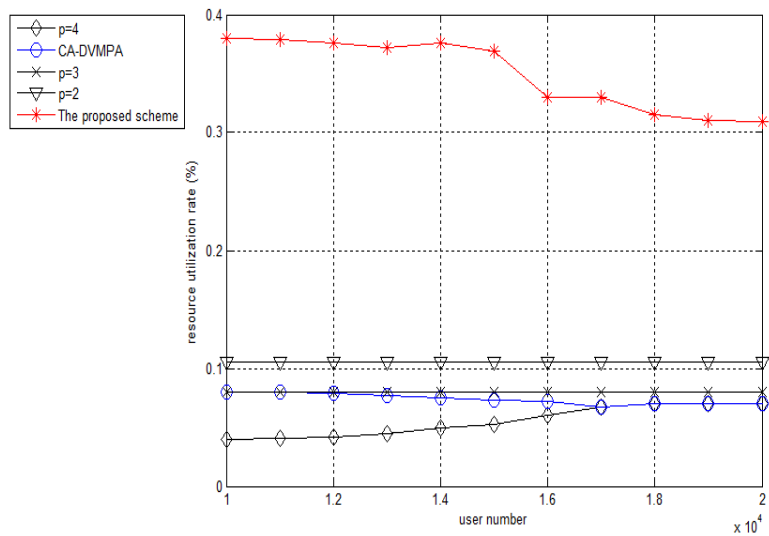


Figure 5. The Resource Utilization Rate Comparison

5. Conclusions

This paper designed a resource scheduling method which has high energy utilization for mobile Internet cloud computing, search method according to the optimization of cloud computing, using improved PSO algorithm, convergence of the time can quickly be reached, so both energy consumption and time delay two aspects.

References

- [1] C. Ge, Z. Sun and N. Wang, "A Survey of Power-Saving Techniques on Data Centers and Content Delivery Networks[J]", *IEEE Communications Surveys & Tutorials*, vol. 15, no. 3, (2013), pp. 1334-1354.
- [2] F-M Liu, P. Shu, H. Jin, L-J Ding, J. Yu, D. Niu and B. Li, "Gearing resource-poor mobile devices with powerful clouds: architectures, challenges, and applications[J]", *IEEE Wireless Communications*, vol. 20, no. 3, (2013), pp. 334 -1354.
- [3] R. Kaewpuang, D. Niyato, P. Wang, *et al.* "A Framework for Cooperative Resource Management in Mobile Cloud Computing[J]", *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 12, (2013), pp. 2685-2700.
- [4] S. Chaisiri, B. Lee, D. Niyato, "Optimization of resource provisioning cost in cloud computing[J]", *IEEE Transactions on Services Computing*, vol. 5, no. 2, (2012), pp. 164-177.
- [5] F. Tian, K. Chen, "Towards Optimal Resource Provisioning for Running MapReduce Programs in Public Clouds[C]", // *IEEE International Conference on Cloud Computing (CLOUD)*, (2011).
- [6] A. Quiroz, H. Kim, M. Parashar, *et al.* "Towards autonomic workload provisioning for enterprise grids and clouds [C]", //2009 10th IEEE/ACM International Conference on Grid Computing, (2009).
- [7] N. Chohan, C. Castillo, M. Spreitzer, *et al.*, "See spot run: using spot instances for mapreduce workflows[C]", //*Proceedings of the 2nd USENIX conference on Hot topics in cloud computing*, USENIX Association, (2010).
- [8] S. Zaman, D. Grosu, "Combinatorial auction-based dynamic vm provisioning and allocation in clouds[C]", //2011 IEEE Third International Conference on Cloud Computing Technology and Science (CloudCom), (2011).
- [9] S. Zaman, D. Grosu, "Combinatorial auction-based allocation of virtual machine instances in clouds[J]", *Journal of Parallel and Distributed Computing*, vol. 73, no. 4, (2013), pp. 495-508.
- [10] A. Kansal, F. Zhao, J. Liu, *et al.* "Virtual machine power metering and provisioning[C]", // *Proceedings of the 1st ACM symposium on Cloud computing*, (2010).
- [11] A. Das, D. Grosu, "Combinatorial auction-based protocols for resource allocation in grids[C]", // *Proceedings 19th IEEE International Parallel and Distributed Processing Symposium*, (2005).

Authors



Yong Lin. Yong Lin received the Bachelor degree in Computer Engineering from Qingdao university of science and technology and Master's degree in Computer Engineering from China Ocean University, CHINA in 2003 and 2011 respectively. He is teaching and researching Computer software at Qingdao Vocational and Technical College of Hotel Management.