# Grid Independent Task Scheduling Multi-Objective Optimization Model and Genetic Algorithm

Hai Zhu
School of Computer Science and Technology Xidian University, Xi'an, China
Email: zhu_sea@163.com

Yuping Wang, Lei Fan and Xiaoli Wang
School of Computer Science and Technology Xidian University, Xi'an, China
Email: {ywang@xidian.end.cn, lfan@mail.xidian.edu.cn, wangxiaolibox@foxmail.com}

***Abstract*—The characteristic of heterogeneous grid environment results in that the task scheduling is constrained by a number of factors such as the length of scheduling, the performance of security, and the cost of scheduling, etc. Firstly, under the consideration of the task scheduling demand for secure performance. A security benefit function and the corresponding time cost model are constructed. Meanwhile, according to the history behavior of grid resource nodes, a simple and efficient node's credibility dynamic evaluation model is built by using the weighted function. Based on these, a new constrained multi-objective grid task scheduling mode is proposed. In order to solve this model, the relationship between the characteristic of the task and the different performance among nodes is used to define the individual relationship matrix; then the subjection degree function is used to transform each relationship matrix into a fuzzy matrix. According to the different effects of each objective in the final objective, each weight in the final decision-making is determined. In doing so, the multi-objective model is transformed into a single-objective model. Through the design of a new mode crossover operator and an uniform mutation operator, a genetic algorithm called MUGA for the transformed problem is proposed and the convergence of proposed algorithm is proved. Simulation results show that the proposed algorithm is better than the compared ones in terms of the length of the task scheduling, security efficiency value, reliability and scheduling costs.**

***Index Terms*—grid computing; tasks scheduling; security benefit function; credibility; genetic algorithm**

## I. INTRODUCTION

Under heterogeneous grid environment, the characteristics of the node resources and the environment are varied. The goal of task scheduling is based on parameters such as the performance of the various resources of the node, communications link state and history behavior to assign different tasks to the corresponding nodes in a reasonable way. When task is implemented on a certain node in a schedule, communications link delays or nodes' performance decrease are likely to increase response time, so the length of the task scheduling is an important factor and needs to be considered in scheduling system. During task implementation in grid platform under scheduling, due to the open environment of grid resource nodes, it is not only have to make sure the results' correctness, but also have to face the improper effects of the network security to the results. Moreover, it is necessary to guarantee that the sensitive data in the execution or transmission will not be leaked, tampered or imposed. So the security of task scheduling is a new factor to be considered [1, 2]. Due to the uncertainty of resources in the grid, tasks implement on a resource node may need re-scheduling or need delays in implementation as a result of node failure attributes. Therefore, tasks' reliability or credibility in grid platform under scheduling is also a factor to be considered [3, 4]. In addition, as a result of grid resource node autonomy, when task assigned to a certain node, in order to encourage node contribute resources, it is required to pay the cost to the node [5], each node in accordance with its own characteristics such as performance and credibility defines its scheduling cost. When the grid user makes the schedule for the tasks, according to its specific needs, he/she may pay more attention to the preference for a shorter scheduling length, or to the consideration of the security and reliability, or to the preference for its own scheduling costs. So, building a scheduling model that can coordinate different goals and meet different users' need is particularly important.

Although traditional task scheduling strategies, such as Min-Min, Max-Min, Sufferage [6, 7], are successfully applied to the scheduling system, but they usually neglect task scheduling needs for security, and also do not take characteristics of grid resources nodes such as open and dynamic environment into consideration, thus it is difficult to make the scheduling system execute effectively in an open, dynamic, real grid environment. In this point of view, for the tasks scheduling problems under heterogeneous grid environment, this paper

Supported by the National Natural Science Foundation of China (No. 60873099), and the National Research Foundation for the Doctoral Program of Higher Education of China(No.20090203110005).


doi:10.4304/jcp.5.12.1907-1915

considered the requirements of tasks to the security demand such as confidentiality, integrity and authenticity when tasks are in implementation or data is in transmission, based on encryption algorithm adopted by the grid resources nodes, Hash Function and authentication technology, a security benefit function and the corresponding time cost model are designed; Meanwhile, in order to ensure that the tasks can be scheduled credibly in the Grid platform, according to the history behavior of grid resource nodes (include the recent successful implementation time of the task, feedback from the previous results of the implementation and resources node's accumulated contribution utilization), a simple and efficient node's credibility evaluation model is built by using the weighted function. Based on these, combined with the length of the task scheduling and scheduling costs, a new constrained multi-objective grid task scheduling model is proposed.

Since different objective functions may be independent of each other, and even conflict, multi-objective constrained grid task scheduling problem has been proved to be an NP hard problem[8], it is difficult to obtain the optimal solution. The current Grid Task Scheduling Algorithm is mainly based on the demand of QoS, using scheduling heuristic to construct the ultimate scheduling structure directly. Reference[9] proposed a heuristic scheduling algorithm which is based on the QoS-Min-min. The task of scheduling is divided into two kinds according to the high or low service quality requirements, and the tasks with the high quality service requirements are arranged first in oeder to meet the needs of users. Reference [10] chooses the least busy node as the executing node each time, and also considers the node load status, but it does not consider the efficiency of the implementation of different nodes, thus it can not be reduced the scheduling length to the minimum. Reference [11] proposes a fuzzy clustering-based heterogeneous grid task scheduling algorithm, which uses fuzzy clustering method to narrow the search space, reduce the unit selection time spent during task scheduling processing, but it does not take into account the cost of scheduling. These heuristic scheduling algorithms are simple and effective, and they have been widely studied and used, but there are also a number of inherent flaws themselves. It is difficult to meet the demands of the grid environment users for a number of objectives at the same time. By considering these shortcomings, this paper uses the relationship between the characteristic of the task and the different performance among nodes to define the individual relationship matrix; then use the subjection degree function to transform each relationship matrix into a fuzzy matrix. By controlling the different parameters for each objective, the task scheduling model can be transformed into single-objective model. Based on this, through the design of a new mode crossover operator and a uniform mutation operator, a genetic algorithm MUGA to solve the problem is proposed. By adjusting the weights factors of different objectives to coordinate the demand trends of different users, so that the scheduling can meet different grid users' demand to different objective optimal values.

## II. SECURITY BENEFIT FUNCTION AND CREDIBILITY MODEL

In grid scheduling system, security and credibility are two different concepts[12], security means that the grid node itself is of inherent safety calculation service guarantee attributes, but credibility is an indicator of resources node's external reliability behavior.

### A. Security benefit function

In general, the main consideration in this section is the confidentiality, integrity and authenticity that generally involved in the grid security. A security benefit function and the corresponding time cost model are constructed.

The guarantee of the security of task scheduling mainly depends on algorithm or method adopted by grid resources node: such as the confidentiality normally relies on the different encryption algorithms (such as IDEA, DES, RC5, Blowfish and RC4, etc.), preventing tasks exposing to other unauthorized users or processes during implementation in grid environment or transmission, so as to avoid disclosure of information. Integrity mainly relies on different Hash function (such as MD4, RIPEMD, SHA-1 and Tiger, etc.), to ensure that the tasks will not be modified or tampered by illegal users during the implementation in grid platform. Authenticity uses a different digital signatures technology (e.g., HMAC-MD5, HMAC-SHA-1 and CBC-MAC-AES, etc.) to identify the authenticity of the communication entities, to ensure that the tasks are submitted by legitimate users.

TABLE I.
THE ENCRYPTION ALGORITHMS AND ITS SECURITY LEVEL VALUE

| **encryption algorithms** | **Algorithm performance (KB/ms)** | **security level value** |
|---|---|---|
| **SEAL** | **168.75** | **0.08** |
| **RC4** | **96.43** | **0.14** |
| **Blowfish** | **37.5** | **0.36** |
| **Knufu/Khafre** | **33.75** | **0.40** |
| **RC5** | **29.35** | **0.46** |
| **Rijndael** | **21.09** | **0.64** |
| **DES** | **15** | **0.90** |
| **IDEA** | **13.5** | **1.00** |

In the following, we take the confidentiality as an example to show how to design the security level value of grid resource node rationally, as well as the corresponding time cost.

Assumed that all the grid resources nodes used eight encryption algorithms in total, as shown in table I.

Let the security level value of the confidentiality of IDEA encryption algorithm corresponding to the lowest performance be 1, then the encryption algorithm's security level value of any other grid resources node $p_j$ based on their performance can be calculated by

$$SL_j^e = 13.5 / \mu_j^e(k), \qquad 1 \le k \le 8 \qquad (1)$$

In this way, based on the different performance of encryption algorithms, set their security values between 0.08 to 1. We can see that the encryption algorithm's security level value of confidentiality is inversely proportion to the performance, the lower the security level value, the higher the performance in general. This is also consistent with the general logic of thinking, under the same conditions, no one would choose a encryption mechanism with low level of security and performance.

Similarly, for grid resource node $p_j$ , based on the Hash function and the type of digital signature technology used, its integrity and authenticity level values $SL_j^e$ and $SL_j^a$ can be obtained respectively. Then integrated security functions of any other grid resources node $p_j$ can be gotten as follows:

$$\begin{aligned} SL_j &= \omega_1 SL_j^e + \omega_2 SL_j^g + \omega_3 SL_j^a \\ s.t \quad & \sum_{i=1}^{3} \omega_i = 1 ,\ \omega_i \ge 0 \\ & 0 \le SL^k \le 1 ,\ k \in \{e, g, a\} \end{aligned} \tag{2}$$

In (2) $\omega_i$ stands for the different weight of security services, the greater the weight, the more important the corresponding security services.

However, at the same time, the security gurantee of task scheduling means the increase of the waiting time, thus it is very important to quantify the time cost for different security level. The security performance time spent for implementation of task $t_i$ on heterogeneous grid resource node $p_j$ is mainly determind by the algorithm type, the amount of data that need to be addressed and the machine's performance. For example, task $t_i$ is implemented on node $p_j$ and we can obtain the Eencryption security value $SL^e(i,j)$, and the performance of encryption algorithm corresponding to its security value is $P(SL_j^e)$, as well as the amount of date need to be encrypted for task $t_i$ is $l_i$, then its corresponding time cost $Costtime(SL^e(i,j))$ can be gotten by

$$Costtime(SL^e(i,j)) = l_i / P(SL_j^e) = l_i / \mu_j^e(k) \ ,\ 1 \le k \le 8 \tag{3}$$

Similarly, the the integrity and authenticity time costs spent for implementation of task $t_i$ on the grid resources node $p_j$ are respectively denoted as $Costtime\ (SL^g(i,j))$ and $Costtime\ (SL^a(i,j))$, then the total time for security effective value model is:

$$f(i,j) = Costtime(SL(i,j)) = \sum_{k \in \{a,e,g\}} Costtime(SL^k(i\ j)) \tag{4}$$

### B. Credibility Model

In heterogeneous grid environment scheduling system, task scheduler evaluates the credibility of each node according to the history behavior of grid resource node (include the recent successful implementation time of the task, feedback from the previous results of the implementation and resources node's accumulated contribution utilization). Each node has its corresponding degree of credibility, and its value mainly show its behavior performance in the past, reflecting the reliability of the node.

The credibility of the node degree is a dynamic function of time, and it is not only relevant to the accumulated usage of grid resource node, but also relevant to the evaluation of credibility that node obtained from the cumulative implementation of tasks. The credibility model of node $p_j$ in the gridan system is defined as follow:

$$\begin{aligned} RL_j &= \alpha \times e^{-TB_j} + \beta \times RS_j + \gamma \times RU_j \\ s.t. \quad & 0 < \alpha, \beta, \gamma < 1 \\ & \alpha + \beta + \gamma = 1 \end{aligned} \tag{5}$$

In (5), $\alpha$ is time reduce weight coefficient, $\beta$ is the credibility evaluation weight coefficient that grid resource node $p_j$ obtained from the cumulative implementation of tasks. $\gamma$ is the resources accumulated utilization weight coefficient. $TB_j$ is the time interval that grid node $p_j$ last time successfully implements till so far. $TB_j > 0$ ; $RU_j$ is accumulated utilization of node $p_j$ , expressed by the ratio between accumulated implementation time and accumulated active time of availible node $p_j$ ; $RS_j$ is the evaluation of credibility that node obtained from the cumulative implementation of tasks, including three factors, namely, time, security and cost. Its calculation model is as follows:

$$RS_j = \chi \times \frac{\sum_i (b_{io} - b_i)}{\sum_i b_{io}} + \delta \times \frac{\sum_i (s_i - s_{io})}{\sum_i s_{io}} + \eta \times \frac{\sum_i (t_{io} - t_i)}{\sum_i t_{io}} \tag{6}$$

In (6), coefficients $\chi$ 、 $\delta$ and $\eta$ are the credibility evaluations that respectively are contributed by cost, security and time, $0 < \chi, \delta, \eta < 1$ and $\chi + \delta + \eta = 1$ 。 $b_{io}$ is the cost of the estimation in which task $t_i$ is implemented on grid node $p_j$, $b_i$ is the actual cost value. Similarly， $s_{io}$ is the security value declaimed by resource node $p_j$, and $s_i$ is the actually obtained security value. $t_{io}$ is the expected finish time， and $t_i$ is the actual finish time. It is clear that when task $t_i$ is implemented on assigned node $p_j$ , if the shorter the time spent, the higher the security value obtained, the less the cost, then the higher the evaluation for node $p_j$ will be obtained. Thus the contribution of node $p_j$ to the ultimate credibility will be greater.

## III. Multi-objective Tasks Scheduling Model and Objective Processing

### A. Problem Description and Modeling

A grid scheduling system is composed of tasks scheduler $R$ , heterogeneous grid resource node sets

$P = \{ p_1, p_2 \cdots p_m \}$ and the composition of the task set $T = \{ t_1, t_2 \cdots t_n \}$. Because we consider heterogeneous system, the estimated task execution time can be expressed in a $n*m$ matrix $E$, in which every element $e_{i,j}$ stands for the estimated execution time for task $t_i$ on node $p_j$. Communication links between nodes can be expressed by a $m*m$ matrix $B$, in which every element $b_{i,j}$ stands for the data transfer rate for task $t_i$ on node $p_j$. The goal of grid task scheduling is to reasonably assign the tasks set $T$ on grid nodes set $P$, so that the scheduling length would be the shortest, the security effective value would be the largest, the reliability would be highest and the cost of the scheduling lowest. Then, based on the security effective function and credibility model in the previous section, an multi-objective constrained task scheduling mathematical model can be expressed as follows:

$$\begin{cases} \min(z_1) = \sum_{i=1}^{n}\sum_{j=1}^{m} x_{ij} C_{ij} \\ \max(z_2) = \sum_{i=1}^{n}\sum_{j=1}^{m} x_{ij} S_{ij} \\ \max(z_3) = \sum_{i=1}^{n}\sum_{j=1}^{m} x_{ij} R_{ij} \\ \min(z_4) = \sum_{i=1}^{n}\sum_{j=1}^{m} x_{ij} Q_{ij} \end{cases} \quad (7)$$

In (7), $z_1$ is grid tasks scheduling length, $z_2$ is the total security effective value, $z_3$ is credibility value obtained by task scheduling, $z_4$ is the cost of scheduling, and $x_{ij}$ is decision variable where its definition is as follow:

$$x_{ij} = \begin{cases} 1 & task\ t_i\ assigned\ on\ node\ p_j \\ 0 & else \end{cases} \quad (8)$$

$S_{ij}$ and $R_{ij}$ respectively indicates the integrated security effective value obtained by task $t_i$ on node $p_j$ and credibility value, $Q_{ij}$ indicates the cost need to be paid by task $t_i$ on node $p_j$. $C_{ij}$ indicates the scheduling length that task $t_i$ is assigned on grid resource node $p_j$. Since the node contains the information including time $a_j$ 、communication time $b_{ij}$ 、task implement time $e_{ij}$ and the cost time to obtain security effectiveness value $f_{ij}$, $C_{ij}$ can be calculated as following:

$$\begin{aligned} C_{ij} &= a_j + b_{ij} + e_{ij} + f_{ij} \\ &= a_j + \left(b_{ij}^0 + mt_i / b_{ij}^1\right) + e_{ij} + f_{ij} \end{aligned} \quad (9)$$

In which communication time $b_{ij}$ includes linking time among nodes $b_{ij}^0$ and data transfer time $mt_i / b_{ij}^1$.

*B. subjects processing*

For multi-objective scheduling system, due to the different evaluation indicators for each objective and different requirements to each subjective according to different scheduling tasks, In this section, the subjection degree function is used to deal with all objectives, and the different weights are assigned to different objectives according to their importance, then the weighted sum of all objective functions is adopted as the new objective.

In the grid task scheduling system in this paper, scheduling pair which is composed of the task set $T$ and the grid resources node set $P$ forms a correlation matrix with respect to some objective $k$ ( $k = 1,2,3,4$ ) (if the system is only concerned about part of the objective, only the corresponding part of the matrix is taken into account) is defined as:

$$X_k = \begin{bmatrix} x_{11}^k & x_{12}^k & \cdots & x_{1m}^k \\ x_{21}^k & x_{22}^k & \cdots & x_{2m}^k \\ \vdots & \vdots & & \vdots \\ x_{n1}^k & x_{n2}^k & \cdots & x_{nm}^k \end{bmatrix} = \left(x_{ij}^k\right)_{n \times m} \quad (10)$$

For any objective in the correlation matrix (10), we use subjection degree function to process it in order to make all objective comparable.

The shorter the scheduling tasks' length is, the better the scheduling. The length of the scheduling task is a indicators of income, and the subjection degree function is defined as follows:

$$y_{ij}^1 = \begin{cases} 0, & x_{ij}^1 < x_{\min}^1 \\ \dfrac{x_{ij}^1 - x_{\min}^1}{x_{\max}^1 - x_{\min}^1}, & x_{\min}^1 \le x_{ij}^1 \le x_{\max}^1 \\ 1, & x_{ij}^1 > x_{\max}^1 \end{cases} \quad (11)$$

In (11), $x_{\min}^1$ and $x_{\max}^1$ are respectively as the maximum value and minimum value of scheduling tasks' length in the correlation matrix $X_1$ 。

The higher the security obtained by the scheduling tasks is, the better the scheduling will be. Because security effectiveness function has already normalized its value to be a real number between 0 and 1, in order to deal with all objectives by minimization, a simple subjection degree function is adopted as follows:

$$y_{ij}^2 = 1 - x_{ij}^2 \quad , \quad 0 \le x_{ij}^2 \le 1 \quad (12)$$

In (12), $x_{ij}^2$ stands for the integrated security value obtained by task $t_i$ being implemented on node $p_j$. Similarly, the greater credibility value that the scheduling tasks obtained, the better the credibility of task scheduling. Thus, maximizing objective $x_{ij}^3$ is changed into minimizing objective $y_{ij}^3$.

The reduction of task scheduling cost and objective satisfaction increase can be considered as a direct proportion relationship, and the following subjection degree formula can be adopted:

$$y_{ij}^4 = 1 - \frac{x_{\min}^4}{x_{ij}} \qquad (13)$$

In (13), $x_{\min}^4$ is the minimum scheduling cost in correlation matrix $X_4$.

According to the processing of subjection degree in (11)~(13)，correlation matrix $X_k$ can be changed into Fuzzy relationship matrix $Y_k$ for objective $k$ ：

$$Y_k = \begin{bmatrix} y_{11}^k & y_{12}^k & \cdots & y_{1m}^k \\ y_{21}^k & y_{22}^k & \cdots & y_{2m}^k \\ \vdots & \vdots & & \vdots \\ y_{n1}^k & y_{n2}^k & \cdots & y_{nm}^k \end{bmatrix} = \left(y_{ij}^k\right)_{n\times m} \qquad (14)$$

After the integration by subbjection degree function for all objectives, the multi-objective task scheduling model can be transformed into the following single-objective optimization model by using matrix (14) as follows:

$$\begin{aligned} \min\left(z^{'}\right) &= \sum_{i=1}^{n}\sum_{j=1}^{m}\left[\omega_1 C_{ij} + \omega_2\left(1-S_{ij}\right) + \omega_3\left(1-R_{ij}\right) + \omega_4 Q_{ij}\right] \\ &= \sum_{i=1}^{n}\sum_{j=1}^{m}\omega_k y_{ij}^k \end{aligned} \qquad (15)$$

In (15) , $0 \le \omega_k \le 1$ , $k$=1, 2, 3, 4 , $\sum_{k=1}^{4}\omega_k = 1$ . According to the task characteristics and different requirements, one can adjust the weights $\omega_k\left(k = 1, 2, 3, 4\right)$ ratio and re-optimize the objective to get a satisfied solution.

## IV. Algorithm Design and Analysis of Convergence

Since multi-objective task scheduling is a NP-hard problem, and genetic algorithms [13] based on evolutionary theory is well suited to solve such complex optimization problems. The mode crossover operator, the even mutation operator, the proposed algorithm and its convergence analysis are given below.

### A. Mode crossover operator

Crossover operator is the most important search operator in GA, in order to enhance the breadth of search, a new crossover operator has been designed, that is, the corresponding bit value of two father individuals are summed up and the result is taken the modulus $m$ , then its offspring is gotten. For example, if $n$=10， $m$=4 , crossover-process is shown in following Figure 1：

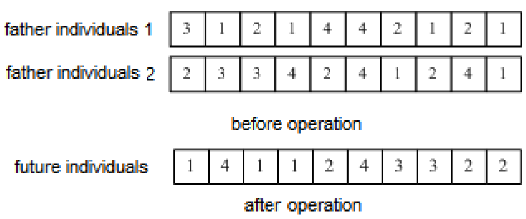



Fig.1 Crossover operator diagram before and after operation

### B. Even mutation operator

Mutation operator is an auxiliary operator, it is generally a small perturbation to the parent individual in order to increase the diversity of the population and to help the algorithm to jump out of the local optimal solution. In this paper, we designed a uniform mutation operator, that is, suppose that $a = \left(a_1 a_2 \cdots a_n\right)$ is the parent individual that participates in mutation, then randomly generated a $x_i \in [1, m]$ as a replacement of $a_i$ from $[1, m]$ for each component of $a$ with a equal probability, and get a new individual $x = \left(x_1 x_2 \cdots x_n\right)$ as the offspring of mutation. This mutation operator is helpful to maintain the diversity of the population, it can prevent the algorithm into a local optimal solution effectively, and this mutation operator has the features of simple and easy implementation.

### C. Genetic algorithm framework

Algorithm 1 (MUGA)

1) (Initial population) Determine the appropriate crossover probability $p_c$ , mutation probability $p_m$ and population size $N$， use direct integer encoding scheme [14]， and generate the initial population $X(0)$ randomly. Set the number of evolution generation $t = 0$ .

2) (Crossover) Select a number of crossover parents pair $\left(X^i, X^j\right)$ from $X(t)$ according to crossover probability $p_c$ to take crossover operation, and the set of the offspring of the crossover is denoted as $O_1$ .

3) (Mutation) For each offspring generated in crossover step 2), do the mutation operation according to mutation probability $p_m$ , and the set of offspring of mutation is denoted as $O_2$ .

4) (Selection) Select the best individual up to now by use of the elite preservation strategy, then randomly select $N-1$ individuals from set $X(t) \cup O_1 \cup O_2$ to fill in the next generation population. The total number of selected individuals for next generation population $X(t+1)$ is $N$ . Let $t = t+1$.

5) (Stop Criterion) When the termination condition is met, stop; otherwise, to step 2).

### D. Convergence Analysis

In order to analyze the convergence of MUGA algorithm, first of all, we introduce some important concepts as follows.

**Definition 1.** Assume that $a^* \in S$ corresponds to the global optimal solution of the grid task scheduling model. If

$$\Pr ob\left\{\lim_{t\to\infty} a^* \in X(t)\right\} = 1 \qquad (16)$$

then the genetic algorithm is said to converge to global optimal solutions with probability 1.

**Definition 2.** Individual $b$ is be reachable from individual $a$ through mutation and crossover if

$$P\{MC(a)=b\}>0 \tag{17}$$

in which $MC(a)$ presents the offspring of individual $a$ produced by crossover and mutation is $b$ , and $P\{\bullet\}$ indicates the probability of random event $\{\bullet\}$.

For a genetic algorithm which has a limited search space , T.Bäck[15] has already proved that: if the genetic algorithm satisfies the following two conditions:

a) For any two individuals $a$ and $b$ in the feasible region $S$ , $b$ is reachable from $a$ through mutation or crossover;

b) Population sequence $X(0), X(1), \cdots, X(t), \cdots$ is monotonous, that is, for $\forall t$ , we have

$$\min\{f(x)|x\in X(t)\}\le \min\{f(x)|x\in X(t+1)\} \tag{18}$$

Then the genetic algorithm will converge to global optimal solutions with probability 1.

**Theorem 1.** (Global convergence) The proposed algorithm MUGA converges to the global optimal solutions with probability 1.

Proof: 1) In algorithm MUGA, for any two individuals $a$ and $b$ , $b$ is reachable from $a$ through mutation or crossover;

In fact, the individual $a$ is selected to take part in crossover. Assume $c$ is an offspring produced by $a$ through crossover, then the probability of individual $c$ selected to take part in mutation is $p_m>0$ , thus, the probability of producing $b$ by the mutation and crossover from $a$ is

$$\Pr ob\{MC(a)=b\}\ge p_c * p_m * \Pr ob\{M(c)=b\} \tag{19}$$

Assume that
$c=(c_1 c_2 \cdots c_n)$ , $b=(b_1 b_2 \cdots b_n)$
from mutation operator we can know that the probability of producing $b_i$ from $c_i$ is $\frac{1}{m-1}$ for any component of $c$ .
Therefore,

$$\Pr ob\{M(c)=b\}=\frac{1}{m-1}\cdot\frac{1}{m-1}\cdots\frac{1}{m-1}=\frac{1}{(m-1)^n}>0 \tag{20}$$

And the probability of producing $b$ by the mutation and crossover from $a$ satisfy:

$$\Pr ob\{MC(a)=b\}\ge p_c * p_m * \Pr ob\{M(c)=b\}=\frac{p_c * p_m}{(m-1)^n}>0 \tag{21}$$

That is, for any two individuals $a$ and $b$ in feasible region $S$ , $b$ is reachable from $a$ through mutation or crossover.

2) From the selection strategy of algorithm MUGA, we can know that $X(0), X(1), \cdots, X(t), \cdots$ is monotonous, that is, condition b) is true.

Based on the proof of 1)、2), we know that algorithm MUGA converge to global optimal solutions with probability 1.

## V. Experimental results and analysis

This section first gives simulation environment and parameter setting, and then compare the proposed genetic algorithm MUGA with the classic performance-driven algorithm Min-Min [9], security-driven algorithm TD-Min-Min [4] [16] scheduling algorithm, to verify the effectiveness of the proposed algorithm at the length of the task scheduling, security, efficiency value, credibility , the cost of scheduling and other performance indicators.

### *A. Environment and parameter setting*

In order to evaluate the performance of the proposed algorithm, a grid simulator is designed based on C++ language, and it is composed of Grid-Generator and Task-Generator. Component Grid-Generator is responsible for the simulation of the grid environment. It can randomly generate different properties (Node processing capability, security level, the credibility and the cost of scheduling) of grid nodes and inter-node links with different communications capability, where the performance of nodes is between [10,100], the security level is between [0.01,1], the initial credibility of nodes is between [0.01,1], the failure rate in unit time randomly generated is between [0.0001,0.0015] , the required cost of nodes to complete tasks is between [10,100], the speed of inter-link transmission is between [1,10] Mb / s; Component Task-Generator can randomly generate heterogeneous tasks of different size, where the short task is defined as [10,100] k, the middle task is define as [100k,1M], and the long task is define as [1,10M].

### *B. Results analysis*

This section will carry out comprehensive experiments for evaluating the proposed MUGA from many aspects such as the length of task scheduling, security value, reliability by use of the short task, the middle task and the long task, respectively. The hardware environment for this simulation experiment is a computer with PⅣ 3.0GHz CPU and 1GB memory, and software environment is the Windows XP Professional operating system.

In the experiments, the parameters values in MUGA are as follows: population size is 100, the probability of cross is 0.8, the mutation rate is 0.02 and the largest number of generations is 300. In order to make the simulation environment close to the real grid environment, we expand the scale of the experiments where the number of nodes is expanded to several hundreds and the number of tasks is expanded to several thousands. In this way, the simulation can be more close to the real network environment.

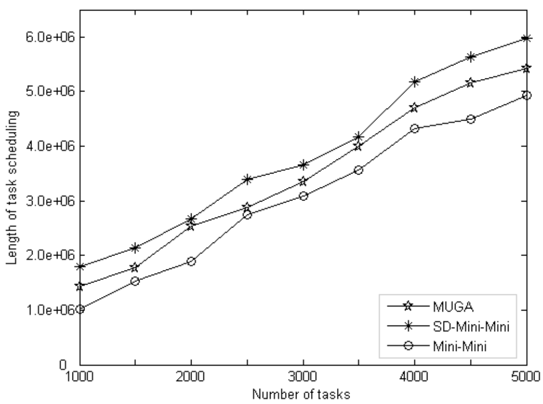
1) The simulation is made and the performance of the proposed algorithm is compared with other algorithms in different performance indicators under the grid stable environment (in the grid systems with m=200 grid nodes and the corresponding number of links) and the change of the number of tasks (sub-tasks number n is from 1000 ~ 5000), (where all the weights are taken the value 0.25, i =1,2,3,4).

From the Fig. 2 (a) we can see that the length of the three tasks scheduling algorithms have increased with the number of tasks increasing, Since Min-Min algorithm do not take the safety factor into consideration, it has the highest success rate; then MUGA follows. This is because algorithm MUGA will take some more time in order to ensure the safety of the tasks, at the same time because MUGA allocates and arranges tasks from the overall perspective, thus it can be of shorter scheduling length.

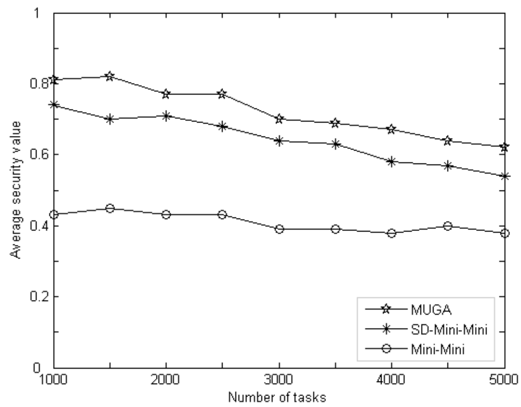
The security benefit function mainly indicates the effectiveness of the security situation of all of tasks in the scheduling. It can be seen from Figure 2(b) that the average security efficiency value for each algorithm has decreased, and MUGA's average value is the largest and decreases fast, while the Min-Min algorithm is without considering safety, the average value of the security is smallest.

value is from the user point of view, referring to the average security efficiency value which was obtained by all successful requested scheduling service. Due to space limitations, this article only tests the latter one. It can be seen from Figure 2 (c) that, under the same situation, MUGA can get a best value of the average of credibility. This is mainly because the credibility model in this paper can improve credibility value of the past nodes of good performance and update credibility value dynamically.

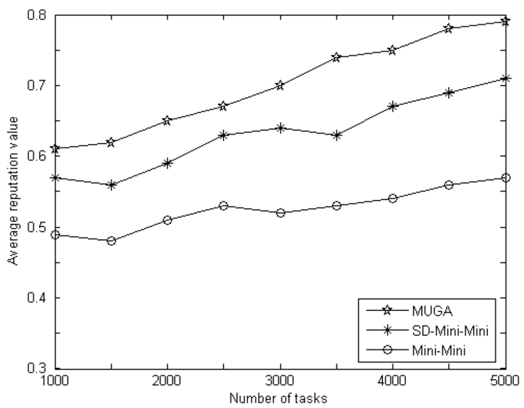
The cost of scheduling mainly refers to the cost the user needs to pay for the node of tasks using, from the user's point of view, for the same task, the less the cost of task scheduling, the better the scheduling. From Figure 2 (d), we can see that as the number of tasks increases, the cost of tasks scheduling increases, under the same situation, because MUGA deal with the problems from the comprehensive details of task scheduling, so it has the less cost in the scheduling.



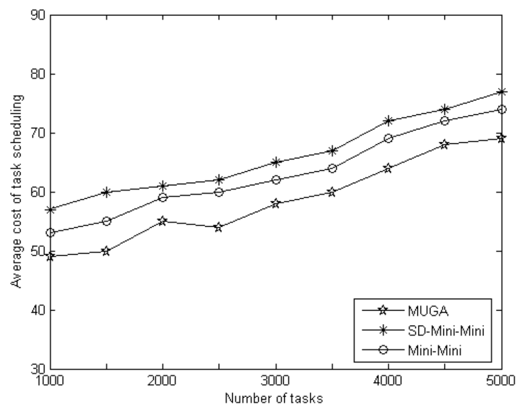

(a) the comparison of tasks scheduling length




(b) the comparison of security effective value




(c) the comparison of credibility value




(d) the comparison of the cost of scheduling

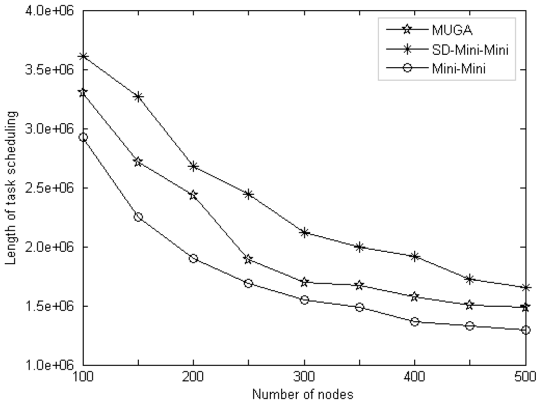
Fig.2 under the condition that nodes number constant, tasks number change, the comparison of different performance indicator

The credibility value obtained by task scheduling mainly indicates the reliability of the implementation of task scheduling. It is relevant to grid resources the tasks are assigned to, and is not related to the implementation order on resources. The total security efficiency value is considered from the systemetic point of view, and it refers to the sum of all maximum security values which were obtained by all successful requested scheduling tasks in the entire system; The average security efficiency

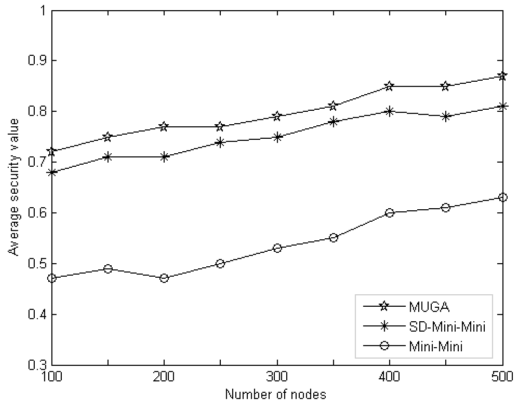
2) The second simulation is to compare the performance of the proposed algorithm with other algorithms in different performance indicators under the changing grid environment (100 ~ 500 grid nodes) and with a fixed number of the tasks (2000 tasks), (where $\omega_i$ is taken as 0.25, i =1, 2, 3, 4).

From the Fig. 3 (a) we can see that the length of the three tasks scheduling algorithms has decreased with the number of nodes increasing. Since Min-Min algorithm do not take the safety factor into consideration, it has the highest success rate; while MUGA allocates tasks from the overall perspective and can has a short scheduling length, and SD-Min-Min of security-driven is the worst. It can be seen from Figure 3 (b) that, with the number of nodes increasing, the security benefit value of the scheduling algorithms which take the security into consideration will increase, and the Min-Min algorithms which do not take the safety factor into consideration will be the worst, and in the same case the MUGA security benefits value is obviously higher than that obtained by other algorithms.

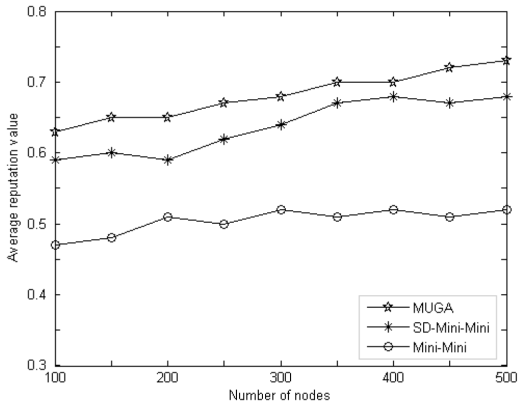
From Figure 3 (c), we can see that, as the nodes increase, MUGA has the most obvious advantage according to the credibility value. This is mainly because the credibility model of this paper dynamically takes the situation of grid task completion into consideration, encourages and improves the past nodes of good performance. From Figure 3 (d), we can see that, as the number of nodes increases, the cost of scheduling decreases. This is mainly because the expansion of the number of nodes can make the scheduling algorithm choose the lower cost of grid nodes priority, and in the same situation, the cost of MUGA is lower than the other scheduling algorithms. This indicates the system has the good extension ability.



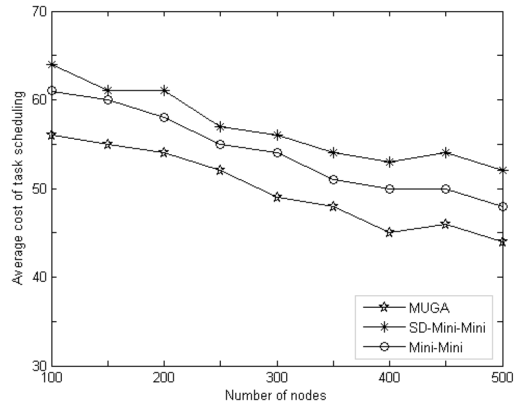
(a) the comparison of tasks scheduling length



(b) the comparison of security effective value



(c) the comparison of credibility value



(d) the comparison of the cost of scheduling

Fig.3 Under the condition that tasks number constant, nodes number change, the comparison of different performance indicator

## VI. CONCLUSIONS

This paper gives a comprehensive analysis of heterogeneous grid task scheduling, according to the features of the grid environment and the different needs of task scheduling, a security benefit function and the credibility dynamic evaluation model are constructed. Based on these, a new model of multi-objective constrained Grid task scheduling is set up. Through the subjection degree function, the multi-objective optimization model is transformed into a single-objective optimization model. By adopting the different weights, the different requirements of the users can be met.

Through the design of new operators, a genetic algorithm called MUGA to solve the problem is proposed and the convergence of this algorithm is proved. Simulation results demonstrate that the proposed algorithm is better than the compared ones in terms of the length of the task scheduling, security efficiency value, reliability and scheduling costs. In the future research, it is necessary to set up new models by taking into account the dynamic needs of different targets, and to develop optimization algorithm for dynamic multi-objective optimization task scheduling.

ACKNOWLEDGMENT

We are grateful for the support of the National Natural Science Foundation of China (No. 60873099), and the National Research Foundation for the Doctoral Program of Higher Education of China (No.20090203110005).

REFERENCES

[1] Xie T, Qin X. Security-aware resource allocation for real-Time parallel jobs on homogenous and heterogeneous clusters [J]. IEEE Trans on Parallel and Distributed Systems, 2008, 19(5): 682-697.
[2] Chakrabarti, A , Damodaran, A , Sengupta, S. Grid computing security: a taxonomy[J]. IEEE Security & Privacy, 2008, 6(1): 44-51.
[3] Song S S, Hwang Kai. Trusted grid computing with security assurance and resource optimization[C]. Proceedings of the 17th International Conference on Parallel and Distributed Computing Systems (PDCS-2004), San Francisco, USA. September, 2004: 15-17.
[4] Song S S , Hwang Kai, Kwok Yu-Kwong. Risk-resilient heuristics and genetic algorithms for security-assured grid job scheduling [J], IEEE Trans. on Computers, 2006, 55(6): 703-719.
[5] Meng Xian-fu, Zhang Xiao-yan. Parallel task scheduling strategy with multi-objective constrains in P2P [J].Computer Integrated Manufacturing Systems, 2008, 14(4): 761-766.
[6] Braun T D, Slegel H J, Becj N. A comparison of eleven static heuristics for mapping a class of independent tasks onto heterogeneous distributed computing systems [J]. Journal of Parallel and Distributed Computing, 2001, 61(6):810-837.
[7] Dogana A, Ozguner F. Scheduling of a meta-task with QoS requirements in heterogeneous computing systems [J]. Journal of Parallel and Distributed Computing, 2006, 66(2): 181-196.
[8] Hironori Kasahara, Seinosuke Narita. Practical multiprocessor scheduling algorithms for efficient parallel Processing [J]. IEEE Trans on Computers, 1984,33(11): 1023-1029.
[9] He X, Sun X , Laszewski G V. QoS guided min-min heuristic for grid task scheduling [J]. Journal of Computer Science and Technology, 2003, 18(4): 442-451.
[10] Dou W, Jia Y, Wang H M. A P2P approach for global computing [C]. Processing of International Parallel and Distributed Processing Symposium. Piscataway, N.J, USA: IEEE Press, 2003: 6-11.
[11] Du X L, Jiang C J, Xu G R. A grid DAG scheduling algorithm based on fuzzy clustering [J]. Journal of Software, 2006, 17(11):2277-2288.
[12] Azzedin F, Maheswaran M. Integrating trust into grid resource management systems[C]. Proceedings of 2002 International Conference on Parallel Processing. Los Alamitos, Cal ,USA: IEEE Computer Society Press, 2002: 47-54.
[13] Zhang W Z, Hu M Z, Zhang H L, et al. A multiobjective evolutionary algorithm for grid Job scheduling of Multi-QoS constraints [J].Journal of Computer Research and Development. 2006，43(11): 1855-1862.
[14] Zhong Y W, Yang J G. Hybid genetic algorithm for independent task scheduling in heterogeneous computing systems [J].Journal of Beijing University of Aeronautics and Astronautics, 2004,30(11):1080-1083.
[15] T.Back. Evolutionary algorithm in theory and practice. Oxford University Press: New York, 1996, 129.
[16] Li K , He Y. ,Liu X. . Security-driven Scheduling Algorithms Based on Eigentrust in Grid[C].Proceeding of the 6th International Conference of Parallel and Distributed Computing Applications and Technologies, Denver, USA, 2005: 1068-1072.

**Hai Zhu** is a Ph.D. candidate of Xidain University of Computer Science and Technology of China. He received the M.S. degree in Computer Science and Technology from Xidian University, Xi'an, China, in 2007. His current research interests include grid computing, tasks scheduling and evolutionary algorithm.

**Yuping Wang** received the M.S. degree in applied mathematics from Xidian University, Xi'an, China, in 1986 and the Ph.D. degree in computational mathematics from Xi'an Jiaotong University,Xi'an, in 1993.He is a Full Professor at Xidian University. He has published more than 100 journal and conference papers. His current research interests include evolutionary computation and optimization theory, network optimization and Data Mining.

**Lei Fan** received the B.S. degree in applied mathematics of Xidian University, Xi'an, China, 2006. He is currently working towards the Ph.D. degree in computer science at Xidian University, Xi'an, China.His current research interests include evolutionary computation, multiobjective optimization and network planning optimization.

**Xiaoli Wang** is a Ph.D. candidate of Xidain University of Computer Science and Technology of China. She received the B.S. degree in Software Engineering from Xidian University, Xi'an, China, in 2008. Her current research interests include evolutionary computation and Data Mining.